(rcsshoe@iastate.edu).
This clone is available through: Biogenetic Services, 801 32nd Ave, Brookings, SD 57006. For further information call 605-697-8500 or contact info@biogeneticservices.com
Seq primer: T7-1 (5' AAT ACG ACT CAC TAT AG 3').

FEATURES Location/Qualifiers
source 1. .689
/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/clone="Gm-r1030-2160"
/lab_host="DH10B"
/clone_lib="Gm-r1030"
/note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This cDNA library was constructed from mRNA isolated from immature cotyledons of greenhouse grown plants (individual seed fresh weight of 100-300mg). The library was prepared using the Life Technologies pSuperScript cDNA library construction kit. Complementary DNA was synthesized from mRNA using a poly(dT) sequence with a NotI restriction site. SalI linkers adapters were ligated to the blunt-ended cDNA fragments followed by NotI digestion. The cDNA fragments were directionally cloned into the NotI-SalI restriction site of the pSPORT1 vector. The ligated cDNA fragments were transformed into E. coli ElectroMax DH10B host cells. This library was constructed by Dr. Lila Vodkin and Dr. Anu Khanna. Note that Gm-r1030 is a re-rack of Gm-c1007."

ORIGIN

Alignment Scores:

Pred. No.:	5.46e-115	Length:	689
Score:	1210.00	Matches:	225
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	29.17%	Indels:	0
DB:	7	Gaps:	0

US-09-419-679-6 (1-757) x CK768972 (1-689)

```
Qy        532 TyrGluLeuLysArgSerTyrAsnTrpLeuGluIleIleAsnProAlaGluThrPheGly 551
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         14 TATGAGCTTAAACGATCTTATAATTGGTTGGAGATAATCAATCCTGCTGAAACTTTTGGT 73

Qy        552 AspIleValIleAspTyrProTyrValGluCysThrSerAlaAlaIleGlnAlaLeuAla 571
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         74 GACATCGTTATTGATTATCCTTATGTGGAATGTACATCAGCAGCGATTCAAGCTTTGGCA 133

Qy        572 SerPheArgLysLeuTyrProGlyHisArgArgGluGluIleGlnHisCysIleAspLys 591
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        134 TCATTTAGGAAATTATATCCTGGGCATCGCCGAGAAGAAATACAACATTGTATCGATAAA 193

Qy        592 AlaThrThrPheIleGluLysIleGlnAlaSerAspGlySerTrpTyrGlySerTrpGly 611
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        194 GCCACTACCTTCATTGAAAAAATACAAGCTTCAGATGGATCATGGTATGGTTCTTGGGGA 253

Qy        612 ValCysPheThrTyrGlyAlaTrpPheGlyValLysGlyLeuIleAlaAlaGlyArgSer 631
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        254 GTTTGCTTCACTTACGGTGCTTGGTTTGGGGTAAAAGGTCTGATTGCTGCTGGAAGGAGT 313

Qy        632 PheSerAsnCysSerSerIleArgLysAlaCysGluPheLeuLeuSerLysGlnLeuPro 651
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        314 TTCAGTAATTGCTCAAGCATCCGTAAAGCTTGTGAATTTCTGCTGTCCAAGCAGCTTCCT 373

Qy        652 SerGlyGlyTrpGlyGluSerTyrLeuSerCysGlnAsnLysValTyrSerAsnLeuGlu 671
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        374 TCTGGTGGCTGGGGAGAGAGTTATCTGTCCTGTCAAAACAAGGTGTATTCAAATCTGGAA 433

Qy        672 GlyAsnArgSerHisValValAsnThrGlyTrpAlaMetLeuAlaLeuIleAspAlaGly 691
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        434 GGCAACAGGTCTCATGTGGTCAACACTGGGTGGGCTATGTTGGCTCTCATTGATGCTGGA 493

Qy        692 GlnAlaLysArgAspSerGlnProLeuHisArgAlaAlaAlaTyrLeuIleAsnSerGln 711
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        494 CAGGCTAAGAGAGATTCGCAACCACTGCACCGGGCAGCTGCATACTTGATAAATTCCCAA 553

Qy        712 LeuGluAspGlyAspProProGlnGlnGluIleMetGlyValPheAsnLysAsnCysMet 731
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        554 TTGGAGGACGGTGACTTTCCGCAGCAGGAAATAATGGGAGTCTTCAACAAGAATTGCATG 613

Qy        732 IleThrTyrAlaAlaTyrArgAsnIleProIleTrpAlaLeuGlyGluTyrGlnSer 751
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        614 ATCACATACGCCGCATACAGAAACATATTCCCCATTTGGGCGTTGGGAGAATACCAATCC 673

Qy        752 GlnValLeuGlnSer 756
              ||||||||||||||
Db        674 CAAGTATTGCAATCT 688
```

Search completed: November 5, 2005, 17:11:16
Job time : 6161 secs

```
Qy        611 GlyValCysPheThrTyrGly 617
              ||||||||||||||||||
Db        782 GCTGTTTGTTTTACCTATGGC 802

RESULT 14
CB651302
LOCUS       CB651302                781 bp    mRNA    linear   EST 08-APR-2003
DEFINITION  OSJNEb16D10.f OSJNEb Oryza sativa (japonica cultivar-group) cDNA
            clone OSJNEb16D10 5', mRNA sequence.
ACCESSION   CB651302
VERSION     CB651302.1  GI:29646295
KEYWORDS    EST.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 781)
  AUTHORS   Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
            Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
  TITLE     Large-scale identification of ESTs involved in the interaction
            between rice and Magnaporthe grisea
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Rod Wing
            Arizona Genomics Institute
            University of Arizona
            Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
            85721-0088, USA
            Tel: 520 626 3967
            Fax: 520 621 9288
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: gta aaa cga cgg cca gtg
            BACKWARD: gga aac agc tat gac cat g
            Plate: 16  row: D  column: 10
            Seq primer: gta aaa cga cgg cca gtg.
FEATURES             Location/Qualifiers
     source          1. .781
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /clone="OSJNEb16D10"
                     /tissue_type="Leaf"
                     /dev_stage="3 week"
                     /lab_host="DH10B"
                     /clone_lib="OSJNEb"
                     /note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
                     XhoI; 24 hrs after innoculation with Rice Blast (Che
                     86061)"
ORIGIN

Alignment Scores:
Pred. No.:              3.76e-116       Length:         781
Score:                  1222.00         Matches:        202
Percent Similarity:     89.96%          Conservative:   31
Best Local Similarity:  77.99%          Mismatches:     26
Query Match:            29.46%          Indels:         0
DB:                     6               Gaps:           0

US-09-419-679-6 (1-757) x CB651302 (1-781)

Qy        142 LeuThrGluGluHisArgLysGluIleCysArgTyrLeuTyrAsnHisGlnAsnLysAsp 161
              |||:::   |||      |||||||||  |||||||||||||||||||||||:::|||
Db          4 TTGTCATCTGAACCTCTGAAGGAGATACGGCGGTATCTCTACAATCATCAGAATGAAGAT 63

Qy        162 GlyGlyTrpGlyLeuHisIleGluGlyProSerThrMetPheGlySerValLeuSerTyr 181
              ||||||||||||||||||||||||||   ||||||||||||||||||   |||:::|||
Db         64 GGAGGATGGGGTTTGCACATTGAGGGCCACAGCACCATGTTTGGCTCAGCGTTGACATAT 123

Qy        182 IleThrLeuArgLeuLeuGlyGluGlyProAsnAspGlyGlnGlyGluMetGluLysAla 201
              ::::::|||||||||||||||||||||||::::   |||   |||   |||||||||
Db        124 GTTTCTTTGAGATTGCTTGGGGAGGGGCCAGATAGCGGCGATGGAGCTATGGAGAAAGGA 183

Qy        202 ArgAspTrpIleLeuGlyHisGlyGlyAlaThrTyrIleThrSerTrpGlyLysMetTrp 221
              |||   |||||||||  ||||||||||||||||||||||||||||||||||    |||
Db        184 CGAAAATGGATTTTAGACCATGGGGGAGCAACTTATATTACATCATGGGGCAAGTTTTGG 243

Qy        222 LeuSerValLeuGlyValTyrGluTrpSerGlyAsnAsnProLeuProProGluIleTrp 241
              |||||||||||||||||:::::|||||||||||||||||||:::|||||||||||||||
Db        244 CTTTCGGTGCTTGGTGTATTTGACTGGTCTGGCAACAACCCAGTGCCACCAGAAATATGG 303

Qy        242 LeuLeuProTyrMetLeuProPheHisProGlyArgMetTrpCysHisCysArgMetVal 261
              ||||||||||||   ||||||  ||||||||||||||||||||||||||||||||||||
Db        304 TTGTTGCCATATTTCCTGCCGATTCATCCAGGGCGAATGTGGTGTCATTGCCGGATGGTT 363

Qy        262 TyrLeuProMetSerTyrLeuTyrGlyLysArgPheValGlyProIleSerProThrVal 281
              ||||||||||||   |||:::|||||||||||||||||||||||:::::|||   :::
Db        364 TATTTGCCTATGTGTTACATTTATGGAAAGAGGTTTGTGGGCCCAGTTACACCAATTATA 423

Qy        282 LeuSerLeuArgLysGluLeuTyrThrValProTyrHisAspIleAspTrpAspGlnAla 301
              |||   ||||||||||||||||||   |||||||||:::::::::|||||||||:::|||
Db        424 TTGGAATTAAGAAAGGAACTCTACGAAGTACCCTACAATGAAGTTGATTGGGACAAGGCT 483

Qy        302 ArgAsnLeuCysAlaLysGluAspLeuTyrTyrProHisProLeuValGlnAspIleLeu 321
              ||||||||||||||||||||||||||||||||||||||||||||   |||||||:::|||
Db        484 CGCAATCTATGTGCTAAGGAAGATCTGTACTATCCACATCCATTCGTGCAGGATGTATTA 543

Qy        322 TrpAlaSerLeuHisLysPheLeuGluProIleLeuMetHisTrpProGlyLysArgLeu 341
              ||||||:::||||||||||||:::||||||   ::::::   |||||||   :::|||
Db        544 TGGGCCACTCTCCACAAATTTGTTGAACCAGCTATGTTGCGTTGGCCTGGGAACAAATTG 603

Qy        342 ArgGluLysAlaIleIleSerAlaLeuGluHisIleHisTyrGluAspGluAsnThrArg 361
              |||||||||||:::   :::   ::::::||||||||||||||||||||||||||||||
Db        604 AGGGAGAAAGCTTTGGACACTGTCATGCAGCATATTCATTATGAAGATGAGAACACCCGA 663

Qy        362 TyrIleCysIleGlyProValAsnLysValLeuAsnMetLeuCysCysTrpValGluAsp 381
              ||||||||||||||||||||||||||||||||||||||||||||   ||||||:::||||||
Db        664 TATATTTGCATTGGTCCAGTAAACAAGGTATTAAATATGCTTGCTTGCTGGATTGAAGAT 723

Qy        382 ProAsnSerGluAlaPheLysLeuHisLeuProArgIleTyrAspTyrLeuTrpIle 400
              |||||||||||||||||||||||||:::||||||:::::|||||||||||||||
Db        724 CCAAACTCAGAGGCATTCAAACTCCACATTCCAAGAGTCCACGATTACCTATGGATT 780

RESULT 15
CK768972
LOCUS       CK768972                689 bp    mRNA    linear   EST 20-FEB-2004
DEFINITION  Gm-r1030-2160 Gm-r1030 Glycine max cDNA clone Gm-r1030-2160 5',
            mRNA sequence.
ACCESSION   CK768972
VERSION     CK768972.1  GI:42723073
KEYWORDS    EST.
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 689)
  AUTHORS   Periappuram,C.C., Li,L., Wurtele,E.S., Westgate,M.E. and
            Nikolau,B.J.
  TITLE     Molecular determination of soybean composition
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Basil J. Nikolau and Cyril C. Periappuram
            Department of Biochemistry, Biophysics and Molecular Biology
            Iowa State University
            2210 Molecular Biology Building, Ames, IA 50011, USA
            Tel: 515 294 9423
            Fax: 515 294 0453
            Email: dimmas@iastate.edu
            Individual base call and confidence value were assigned using the
            Phred software (http://www.phrap.org/). Overall sequence quality
            assessment and vector trimming were conducted using the Lucy
            software (http://www.tigr.org/software/).
            This clone was originally generated by the Public Soybean EST
            Project (http://129.186.26.94/soybeanest.html)/Shoemaker, R
```

```
Qy        189 GluGlyProAsnAspGlyGlnGlyGluMetGluLysAlaArgAspTrpIleLeuGlyHis 208
              |||||||||:::   |||   |||   |||||||||   |||   |||||||||   |||
Db        302 GAGGGGCCAGATAGCGGCGATGGAGCTATGGAGAAAGGACGAAAATGGATTTTAGACCAT 361

Qy        209 GlyGlyAlaThrTyrIleThrSerTrpGlyLysMetTrpLeuSerValLeuGlyValTyr 228
              |||||||||||||||||||||||||||||||   ||||||||||||||||||||:::
Db        362 GGGGGAGCAACTTATATTACATCATGGGGCAAGTTTTGGCTTTCGGTGCTTGGTGTATTT 421

Qy        229 GluTrpSerGlyAsnAsnProLeuProProGluIleTrpLeuLeuProTyrMetLeuPro 248
              :::|||||||||||||||||:::|||||||||||||||||||||||||   |||||||
Db        422 GACTGGTCTGGCAACAACCCAGTGCCACCAGAAATATGGTTGTTGCCATATTTCCTGCCG 481

Qy        249 PheHisProGlyArgMetTrpCysHisCysArgMetValTyrLeuProMetSerTyrLeu 268
                 |||||||||||||||||||||||||||||||||||||||||||||||||   |||:::
Db        482 ATTCATCCAGGGCGAATGTGGTGTCATTGCCGGATGGTTTATTTGCCTATGTGTTACATT 541

Qy        269 TyrGlyLysArgPheValGlyProIleSerProThrValLeuSerLeuArgLysGluLeu 288
              |||||||||||||||||||||:::::|||   :::|||   ||||||||||||||||
Db        542 TATGGAAAGAGGTTTGTGGGCCCAGTTACACCAATTATATTGGAATTAAGAAAGGAACTC 601

Qy        289 TyrThrValProTyrHisAspIleAspTrpAspGlnAlaArgAsnLeuCysAlaLysGlu 308
              |||   |||||||||:::::::::|||||||||:::|||||||||||||||||||||||
Db        602 TACGAAGTACCCTACAATGAAGTTGATTGGGACAAGGCTCGCAATCTATGTGCTAAGGAA 661

Qy        309 AspLeuTyrTyrProHisProLeuValGlnAspIleLeuTrpAlaSerLeuHisLysPhe 328
              |||||||||||||||||||   |||||||||:::|||||||||:::|||||||||||||
Db        662 GATCTGTACTATCCACATCCATTCGTGCAGGATGTATTATGGGCCACTCTCCACAAATTT 721

Qy        329 LeuGluProIleLeuMetHisTrpProGlyLysArgLeuArgGluLysAlaIleIleSer 348
              :::||||||   ::::::   |||||||||   :::|||||||||||||||:::   :::
Db        722 GTTGAACCAGCTATGTTGCGTTGGCCTGGGAACAAATTGAGGGAGAAAGCTTTGGACACT 781

Qy        349 AlaLeuGluHisIleHisTyrGluAspGluAsnThrArgTyrIleCys 364
                 ::::::|||||||||||||||||||||||||||||||||||||
Db        782 GTCATGCAGCATATTCATTATGAAGATGAGAACACCCGATATATTTGC 829

RESULT 13
CB673111
LOCUS       CB673111                803 bp    mRNA    linear   EST 09-APR-2003
DEFINITION  OSJNEe07H20.f OSJNEe Oryza sativa (japonica cultivar-group) cDNA
            clone OSJNEe07H20 5', mRNA sequence.
ACCESSION   CB673111
VERSION     CB673111.1  GI:29676836
KEYWORDS    EST.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 803)
  AUTHORS   Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
            Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
  TITLE     Large-scale identification of ESTs involved in the interaction
            between rice and Mágnaporthe grisea
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Rod Wing
            Arizona Genomics Institute
            University of Arizona
            Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
            85721-0088, USA
            Tel: 520 626 3967
            Fax: 520 621 9288
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: gta aaa cga cgg cca gtg
            BACKWARD: gga aac agc tat gac cat g
            Plate: 07  row: H  column: 20
            Seq primer: gta aaa cga cgg cca gtg.
FEATURES             Location/Qualifiers
     source          1. .803
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /clone="OSJNEe07H20"
                     /tissue_type="Leaf"
                     /dev_stage="3 week"
                     /lab_host="DH10B"
                     /clone_lib="OSJNEe"
                     /note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
                     XhoI; 24 hrs after innoculation with Rice Blast (70-15)"
ORIGIN

Alignment Scores:
Pred. No.:                3.25e-118      Length:          803
Score:                    1242.00        Matches:         215
Percent Similarity:       91.39%         Conservative:    29
Best Local Similarity:    80.52%         Mismatches:      23
Query Match:              29.94%         Indels:          0
DB:                       6              Gaps:            0

US-09-419-679-6 (1-757) x CB673111 (1-803)

Qy        351 GluHisIleHisTyrGluAspGluAsnThrArgTyrIleCysIleGlyProValAsnLys 370
              :::|||||||||||:::||||||||||||||||||||||||||||||||||||||||||
Db          2 CAGCATATTCATTATGACGATGAGAACACCCGATATATTTGCATTGGTCCAGTAAACAAG 61

Qy        371 ValLeuAsnMetLeuCysCysTrpValGluAspProAsnSerGluAlaPheLysLeuHis 390
              ||||||||||||||   |||||:::|||||||||||||||||||||||||||||||||||
Db         62 GTATTAAATATGCTTGCTTGCTGGATTGAAGATCCAAACTCAGAGGCATTCAAACTCCAC 121

Qy        391 LeuProArgIleTyrAspTyrLeuTrpIleAlaGluAspGlyMetLysMetGlnGlyTyr 410
              :::|||||:::::|||||||||||||||||||||||||||||||||||||||||||||
Db        122 ATTCCAAGAGTCCACGATTACCTATGGATTGCAGAAGATGGCATGAAAATGCAGGGTTAT 181

Qy        411 AsnGlySerGlnLeuTrpAspThrAlaPheAlaValGlnAlaIleIleAlaSerAsnLeu 430
              ||||||||||||||||||||||||||||||   ||||||||||:::|||:::   |||
Db        182 AATGGAAGCCAGCTGTGGGACACAGCTTTCACAGTTCAAGCTATAGTGGCTACTGGCCTC 241

Qy        431 IleGluGluPheGlyProThrIleArgLysAlaHisThrTyrIleLysAsnSerGlnVal 450
              ||||||||||||||||||||||:::::   ||||||   |||||||   :::||||||
Db        242 ATTGAAGAATTTGGTCCTACTCTTAAACTAGCACATGGCTACATAAAGAAAACGCAGGTT 301

Qy        451 LeuGluAspCysProGlyAspLeuAsnLysTrpTyrArgHisIleSerLysGlyAlaTrp 470
              ::::::|||||||||||||||||:::::|||||||||||||||||||||||||||||||
Db        302 ATCGATGACTGCCCTGGAGATCTTAGTCAGTGGTACCGCCACATATCTAAAGGTGCATGG 361

Qy        471 ProPheSerThrGlyAspHisGlyTrpProIleSerAspCysThrAlaGluGlyLeuLys 490
              ||||||||||||   ||||||||||||||||||||||||||||||||||||||||||||
Db        362 CCCTTTTCTACTGCTGATCATGGTTGGCCTATATCAGATTGCACTGCAGAAGGACTTAAG 421

Qy        491 AlaValLeuLeuLeuSerLysIleAlaProGluIleValGlyGluProIleAspValLys 510
              |||   |||||||||||||||||:::|||:::|||||||||||   ::::::|||
Db        422 GCGGCATTATTGCTATCGAAGATTTCTCCAGATATTGTTGGCGAAGCAGTGGAAGTTAAT 481

Qy        511 ArgLeuTyrAspSerValAsnValIleLeuSerLeuGlnAsnGluAspGlyGlyPheAla 530
              ||||||||||||||||||   ::::::|||       |||::::::||||||||||||
Db        482 AGACTGTATGATTCTGTCAATTGTTTGATGTCATACATGAATGATAATGGTGGATTTGCA 541

Qy        531 ThrTyrGluLeuLysArgSerTyrAsnTrpLeuGluIleIleAsnProAlaGluThrPhe 550
              ||||||||||||   |||||||||   |||||||||:::||||||||||||||||||||
Db        542 ACATATGAACTCACAAGGTCTTATGCCTGGCTGGAGCTTATCAATCCTGCTGAGACCTTT 601

Qy        551 GlyAspIleValIleAspTyrProTyrValGluCysThrSerAlaAlaIleGlnAlaLeu 570
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        602 GGGGACATTGTGATTGATTATCCTTATGTGGAATGCACTTCAGCAGCAATTCAGGCCCTG 661

Qy        571 AlaSerPheArgLysLeuTyrProGlyHisArgArgGluGluIleGlnHisCysIleAsp 590
                ::|||:::||||||||||||||||||:::   ||||||   :::||||||
Db        662 ACAGCATTTAAAAAGCTCTACCCTGGACACCGCAAGAGTGAAATAGACAACTGTATAAGC 721

Qy        591 LysAlaThrThrPheIleGluLysIleGlnAlaSerAspGlySerTrpTyrGlySerTrp 610
              ||||||   :::||||||||||   ||||||   ||||||||||||||||||||||||||
Db        722 AAAGCTGCTAGCTTTATTGAGGGTATTCAAAAAAGCGATGGTTCATGGTATGGTTCTTGG 781
```

size-selected to contain the 0.5-3 kb size fraction."
ORIGIN

Alignment Scores:

Pred. No.:	9.57e-120	Length:	840
Score:	1257.00	Matches:	221
Percent Similarity:	88.89%	Conservative:	27
Best Local Similarity:	79.21%	Mismatches:	31
Query Match:	30.30%	Indels:	0
DB:	6	Gaps:	0

US-09-419-679-6 (1-757) x CB976421 (1-840)

```
Qy         29 AspProLysLeuGlySerProGlnAspLeuLeuGluIleGluLysAlaArgGlnAsnPhe 48
              ||||||   |||||||||   :::          ||||||||||:::||||||:::   |||
Db          2 GATCCCGATCTGGGATCGGCTGAACAGCGTGAGGAGATCGAGAGAGCTCGTGAGACGTTC 61

Qy         49 HisAspAsnArgPheThrHisLysHisSerAlaAspLeuLeuMetArgMetGlnPheAla 68
                 :::||||||       ||||||||||||||||||:::||||||:::||||||
Db         62 TCGAAGCATCGATTCGAGAAGAAGCATAGTGCAGATCTAATCATGCGTATTCAGTTTTTG 121

Qy         69 ArgGluAsnProThrArgGluValLeuProLysValGlyValLysAspIleGluAspVal 88
              :::|||||||           |||||||||:::|||   |||||||||   |||||||||
Db        122 AAGGAGAATGCTGGTACTGTAGTTTTACCCCAAGTCAAAGTGAAAGATACAGAGGATGTT 181

Qy         89 ThrGlnGluIleValThrLysThrLeuArgArgAlaValSerPheHisSerThrLeuGln 108
              |||::::::   ||||||   ||||||||||||||||:::::|||||||||||||||||
Db        182 ACAGAGGATGCAGTGACAAACACATTAAGAAGGGCTATCAATTTTCATTCAACTCTCCAG 241

Qy        109 CysHisAspGlyHisTrpProGlyAspTyrGlyGlyProMetPheLeuMetProGlyLeu 128
              ||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db        242 GCTCATGATGGACACTGGCCAGGGGATTATGGAGGTCCCATGTTTCTTTTGCCCGGCTTG 301

Qy        129 ValIleThrLeuSerIleThrGlyAlaLeuAsnThrValLeuThrGluGluHisArgLys 148
              :::|||||||||||||||||||   ||||||   ||||||::::::|||||||||:::
Db        302 ATCATTACTCTTTCGATTACTGGGGTGCTGAATGCTGTCTTGTCAAAAGAACATAGACAA 361

Qy        149 GluIleCysArgTyrLeuTyrAsnHisGlnAsnLysAspGlyGlyTrpGlyLeuHisIle 168
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 GAGATGTGCCGTTATCTCTATAACCATCAGAACAAAGATGGTGGGTGGGGTTTACATATC 421

Qy        169 GluGlyProSerThrMetPheGlySerValLeuSerTyrIleThrLeuArgLeuLeuGly 188
              ||||||||||||||||||||||:::||||||:::|||:::||||||||||||||||||
Db        422 GAGGGCCCAAGCACCATGTTTGGTACTGTCTTGAACTATGTAACTTTAAGGTTGCTTGGT 481

Qy        189 GluGlyProAsnAspGlyGlnGlyGluMetGluLysAlaArgAspTrpIleLeuGlyHis 208
              ||||||   ||||||    |||   |||||||||   ||||||||||||   |||
Db        482 GAAGGAGCTAATGATGCAGATGGGGCAATGGAGAAAGGGCGTGACTGGATTCTGAATCAT 541

Qy        209 GlyGlyAlaThrTyrIleThrSerTrpGlyLysMetTrpLeuSerValLeuGlyValTyr 228
              ||||||||||   ||||||||||||||||||||||||||||||||||||||||||:::
Db        542 GGTGGTGCAACTGCAATAACCTCATGGGGAAAAATGTGGCTTTCAGTACTTGGAGTATTT 601

Qy        229 GluTrpSerGlyAsnAsnProLeuProProGluIleTrpLeuLeuProTyrMetLeuPro 248
              ||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        602 GAATGGTCTGGCAATAATCCTCTGCCCCCTGAAATATGGCTTCTTCCCTATATCCTACCA 661

Qy        249 PheHisProGlyArgMetTrpCysHisCysArgMetValTyrLeuProMetSerTyrLeu 268
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        662 GTCCGTCCAGGAAGGATGTGGTGTCACTGCAGGATGGTTTATTTGCCTATGTCATACTTA 721

Qy        269 TyrGlyLysArgPheValGlyProIleSerProThrValLeuSerLeuArgLysGluLeu 288
              ||||||||||||||||||||||:::::|||||||||||||||||||||||||||||||
Db        722 TATGGGAAGAGGTTTGTTGGTCCTCTCACACCTACAGTATTGTCTTTGAGGAAGGAGCTC 781

Qy        289 TyrThrValProTyrHisAspIleAspTrpAspGlnAlaArgAsnLeuCysAlaLys 307
              ||||||||||||||||:::||||||||:::||||||||||||||||||||||||
Db        782 TATACTGTTCCATATCATGAAATCGATTGGAATCAAGCACGAAACCTATGTGCCAAG 838
```

RESULT 12
CB642220
LOCUS CB642220 831 bp mRNA linear EST 08-APR-2003
DEFINITION OSJNEb02E16.f OSJNEb Oryza sativa (japonica cultivar-group) cDNA clone OSJNEb02E16 5', mRNA sequence.
ACCESSION CB642220
VERSION CB642220.1 GI:29637211
KEYWORDS EST.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 831)
AUTHORS Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E., Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
TITLE Large-scale identification of ESTs involved in the interaction between rice and Magnaporthe grisea
JOURNAL Unpublished (2003)
COMMENT Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ 85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu
PCR PRimers
FORWARD: gta aaa cga cgg cca gtg
BACKWARD: gga aac agc tat gac cat g
Plate: 02 row: E column: 16
Seq primer: gta aaa cga cgg cca gtg.
FEATURES Location/Qualifiers
source 1. .831
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="OSJNEb02E16"
/tissue_type="Leaf"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_lib="OSJNEb"
/note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2: XhoI; 24 hrs after innoculation with Rice Blast (Che 86061)"
ORIGIN

Alignment Scores:

Pred. No.:	2.45e-119	Length:	831
Score:	1253.00	Matches:	208
Percent Similarity:	88.41%	Conservative:	36
Best Local Similarity:	75.36%	Mismatches:	32
Query Match:	30.21%	Indels:	0
DB:	6	Gaps:	0

US-09-419-679-6 (1-757) x CB642220 (1-831)

```
Qy         89 ThrGlnGluIleValThrLysThrLeuArgArgAlaValSerPheHisSerThrLeuGln 108
              |||   |||   |||       :::|||:::||||||::::::   :::||||||   |||
Db          2 ACAGGCGAAGCTGTATTATCCTCTCTAAAAAGGGCAATTGCTCGTTATTCTACTTTTCAA 61

Qy        109 CysHisAspGlyHisTrpProGlyAspTyrGlyGlyProMetPheLeuMetProGlyLeu 128
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 GCACATGATGGACACTGGCCTGGGGATTATGGTGGTCCCATGTTCCTTATGCCAGGCTTG 121

Qy        129 ValIleThrLeuSerIleThrGlyAlaLeuAsnThrValLeuThrGluGluHisArgLys 148
              :::|||||||||   ::::::|||||||||||||||   |||:::   ||||||:::|||
Db        122 ATTATAACATTATATGTGAGCGGAGCGCTGAACACTGCTTTGTCATCTGAACATCAGAAG 181

Qy        149 GluIleCysArgTyrLeuTyrAsnHisGlnAsnLysAspGlyGlyTrpGlyLeuHisIle 168
              ||||||   ||||||||||||||||||||||||||:::||||||||||||||||||||
Db        182 GAGATACGGCGGTATCTCTACAATCATCAGAATGAAGATGGAGGATGGGGTTTGCACATT 241

Qy        169 GluGlyProSerThrMetPheGlySerValLeuSerTyrIleThrLeuArgLeuLeuGly 188
              ||||||   ||||||||||||||||||   |||:::|||::::::||||||||||||||
Db        242 GAGGGCCACAGCACCATGTTTGGCTCAGCGTTGACATATGTTTCTTTGAGATTGCTTGGG 301
```

```
  AUTHORS   Lai,J., Dey,N., Kim,C.S., Bharti,A.K., Rudd,S., Mayer,K.F.,
            Larkins,B., Becraft,P. and Messing,J.
  TITLE     Characterization of the maize endosperm transcriptome and its
            comparison to the rice genome
  JOURNAL   Genome Res. 14 (10), 1932-1937 (2004)
COMMENT     Contact: Lai, Jinsheng
            Dr. Joachim Messing's lab
            Waksman Institute, Rutgers University
            190 Frelinghuysen Rd., Piscataway, NJ 08854, USA
            Tel: 732-445-3801
            Fax: 732-445-5735
            Email: jlai@waksman.rutgers.edu
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .850
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /cultivar="W22"
                     /db_xref="taxon:4577"
                     /tissue_type="Endosperm of 7-23DAP"
                     /clone_lib="Endosperm_5"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI"
ORIGIN

Alignment Scores:
Pred. No.:              4.11e-124       Length:         850
Score:                  1299.00         Matches:        220
Percent Similarity:     89.40%          Conservative:   33
Best Local Similarity:  77.74%          Mismatches:     30
Query Match:            31.32%          Indels:         0
DB:                     6               Gaps:           0

US-09-419-679-6 (1-757) x CD440985 (1-850)

Qy        243 LeuProTyrMetLeuProPheHisProGlyArgMetTrpCysHisCysArgMetValTyr 262
              |||||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTGCCATATCTCCTGCCATTTCACCCAGGGAGGATGTGGTGTCATTGTCGAATGGTGTAT 60

Qy        263 LeuProMetSerTyrLeuTyrGlyLysArgPheValGlyProIleSerProThrValLeu 282
              ||||||:::   |||:::|||||||||||||||||||||   |||:::|||   ||||||
Db         61 TTGCCTTTGTGCTACATCTATGGGAAGAGGTTTGTTGGTCGAATCACACCACTCGTGTTG 120

Qy        283 SerLeuArgLysGluLeuTyrThrValProTyrHisAspIleAspTrpAspGlnAlaArg 302
                 ||||||||||||||:::      ||||||      |||||||||||:::||||||
Db        121 GAATTAAGAAAGGAACTTTTCAAAGACCCTTACAGCAAGATTGATTGGGACAAGGCCCGC 180

Qy        303 AsnLeuCysAlaLysGluAspLeuTyrTyrProHisProLeuValGlnAspIleLeuTrp 322
              ||||||||||||||||||||||||||||||||||||||   |||||||||:::||||||
Db        181 AATCTATGTGCTAAGGAAGATCTGTACTATCCACACCCATTTGTTCAAGATGTCTTATGG 240

Qy        323 AlaSerLeuHisLysPheLeuGluProIleLeuMetHisTrpProGlyLysArgLeuArg 342
              |||:::|||||||||||||:::||||||:::::|||   ||||||      :::||||||
Db        241 GCCACTCTCCATAAATTTGTTGAACCAGTTATGATGAGTTGGCCTGTTAGTAAATTGAGG 300

Qy        343 GluLysAlaIleIleSerAlaLeuGluHisIleHisTyrGluAspGluAsnThrArgTyr 362
              |||||||||:::   ::|||:::::|||:::|||||||||||||||||||||||||||||
Db        301 GAGAAAGCTCTGGAGACAGCCATGCAACATGTTCACTATGAGGATGAGAACACTCGGTAC 360

Qy        363 IleCysIleGlyProValAsnLysValLeuAsnMetLeuCysCysTrpValGluAspPro 382
              ||||||||||||||||||||||||||||||||||||||   ||||||:::|||||||||
Db        361 ATTTGCATTGGTCCTGTAAACAAGGTATTAAATATGCTTGCTTGCTGGATTGAAGATCCA 420

Qy        383 AsnSerGluAlaPheLysLeuHisLeuProArgIleTyrAspTyrLeuTrpIleAlaGlu 402
              ||||||||||||||||||||||:::|||||||:::|||||||||||||||:::||||||
Db        421 AACTCAGAGGCCTTTAAACTCCACATCCCACGAGTCTATGATTACTTGTGGCTTGCTGAA 480

Qy        403 AspGlyMetLysMetGlnGlyTyrAsnGlySerGlnLeuTrpAspThrAlaPheAlaVal 422
              ||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db        481 GATGGCATGAAGATGCAGGGCTATAATGGTAGCCAACTTTGGGATACAGCTTTCATAGTT 540

Qy        423 GlnAlaIleIleAlaSerAsnLeuIleGluGluPheGlyProThrIleArgLysAlaHis 442
              |||||||||:::|||:::|||||||   |||||||||||||||||:::::   ||||||

Db        541 CAAGCCATTGTGGCTACCAACCTTACCGAAGAGTTTGGTCCCACCCTTAAACTAGCACAC 600

Qy        443 ThrTyrIleLysAsnSerGlnValLeuGluAspCysProGlyAspLeuAsnLysTrpTyr 462
              |||||||||   ||||||||||||:::|||||||||||||||||||||   ||||||
Db        601 AACTATATCAAGAAATCACAGGTTCTTGACGACTGCCCTGGTGATCTGAATGACTGGTAC 660

Qy        463 ArgHisIleSerLysGlyAlaTrpProPheSerThrGlyAspHisGlyTrpProIleSer 482
              ||||||   ||||||||||||||||||||||||||   ||||||||||||||||||||
Db        661 CGCCACACATCTAAAGGCGCATGGCCATTCTCAACCGCGGATCATGGTTGGCCTATATCT 720

Qy        483 AspCysThrAlaGluGlyLeuLysAlaValLeuLeuLeuSerLysIleAlaProGluIle 502
              ||||||||||||||||||||||||||||   ||||||||||||||||:::|||||||||
Db        721 GATTGCACTGCTGAAGGACTAAAGGCTTCATTATTGTTATCAAAGATTTCTCCTGAAATT 780

Qy        503 ValGlyGluProIleAspValLysArgLeuTyrAspSerValAsnValIleLeuSerLeu 522
              ||||||   ||||||:::      |||   ||||||:::|||:::   ::::::|||
Db        781 GTTGGTGGACCGATAGAAGCTAATAGATTTTATGATGCTGTCAGTTGTCTGATGTCTTAT 840

Qy        523 GlnAsnGlu 525
                 |||:::
Db        841 ATGAATGAT 849

RESULT 11
CB976421
LOCUS       CB976421                840 bp    mRNA    linear   EST 01-MAY-2003
DEFINITION  CAB40002_IIIa_Fa_G04 Cabernet Sauvignon Berry - CAB4 Vitis vinifera
            cDNA clone CAB40002_IIIa_Fa_G04 5', mRNA sequence.
ACCESSION   CB976421
VERSION     CB976421.1  GI:30299627
KEYWORDS    EST.
SOURCE      Vitis vinifera
  ORGANISM  Vitis vinifera
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; Vitaceae; Vitis.
REFERENCE   1  (bases 1 to 840)
  AUTHORS   Goes da Silva,F., Iandolino,A., Lim,H., Baek,J., Leslie,A., Xu,J.,
            Jones,K. and Cook,D.
  TITLE     Expressed sequence tags from Vitis vinifera 'Cabernet sauvignon'
            berries at various developmental stages
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Douglas Cook, PhD
            CAES Genome Facility
            UC Davis, Plant Pathology
            One Shields Ave, Davis, CA 95616, USA
            Tel: 530 754 6561
            Fax: 530 754 6617
            Email: drcook@ucdavis.edu
            Seq primer: ACGGTACCGGACATATGCC.
FEATURES             Location/Qualifiers
     source          1. .840
                     /organism="Vitis vinifera"
                     /mol_type="mRNA"
                     /cultivar="Cabernet Sauvignon"
                     /db_xref="taxon:29760"
                     /clone="CAB40002_IIIa_Fa_G04"
                     /sex="Hermaphrodite"
                     /dev_stage="Berry on stage II, 9 mm"
                     /lab_host="DH5alpha"
                     /clone_lib="Cabernet Sauvignon Berry - CAB4"
                     /note="Organ: Berry; Vector: pDNR; Site_1: SfiI; Site_2:
                     SfiI; CAB4 is a cDNA library of Vitis vinifera cv.
                     'Cabernet Sauvignon' Clone 8 berries. Sampled berries were
                     collected from field-grown vines during stage II of berry
                     growth (berries were green and hard) at approximately 60
                     days after full bloom.The average berry size was 9
                     milimeters. Sampled vines were located at the University
                     of California, Davis, Experimental Vineyard. cDNAs were
                     made by oligo-dT priming and directionally cloned. 5'and
                     3' adaptors were used in cloning as follows:
                     5'-AAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGG-3' and
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)NN-3'. Library was
                     constructed using the Clontech Creator SMART kit and
```

```
Db        601 GGTTCTTGGGGAGTTTGCTTCACTTATGGCACTTGGTTTGGGGTAAAAGGGCTGATTGCT 660

Qy        628 AlaGlyArgSerPheSerAsnCysSerSerIleArgLysAlaCysGluPheLeuLeuSer 647

Db        661 GCTGGGAAGAGTTTCAGTAATTGCTTAAGCATCCGTAAAGCTTGTGACTTTTTGCTGTCG 720

Qy        648 LysGlnLeuProSerGlyGlyTrp-GlyGluSerTyrLeuSerCysGlnAsnLysValTy 667

Db        721 AAGCAGCTCCCATCTGGTGGTTGGGGGAGAGAGTTATCTTTCATGTCAAAACAAGGTGTA 780

Qy        667 rSerAsnLeuGluGlyAsnArgSerHisValValAsnThrGlyTrpAlaMetLeuAlaLe 687

Db        781 TTCAAATTTGGAAAGCCACAGGTCCCATGTAGTAAACACTGGGTGGGCTATGCTGGCTCT 840

Qy        687 uIleAspAlaGlyGlnAla 693

Db        841 CATTGGAGCTGAACAGGCC 859

RESULT 9
CK251308
LOCUS       CK251308                 958 bp    mRNA    linear   EST 30-JUL-2004
DEFINITION  EST734945 potato callus cDNA library, normalized and full-length
            Solanum tuberosum cDNA clone POCBM46 5' end, mRNA sequence.
ACCESSION   CK251308
VERSION     CK251308.1  GI:39804211
KEYWORDS    EST.
SOURCE      Solanum tuberosum (potato)
  ORGANISM  Solanum tuberosum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; lamiids; Solanales; Solanaceae; Solanum.
REFERENCE   1  (bases 1 to 958)
  AUTHORS   Buell,C.R., Hart,A., Zismann,V., Karamycheva,S.A. and Baker,B.
  TITLE     Generation of ESTs from potato callus tissue
  JOURNAL   Unpublished (2003)
COMMENT     Other_ESTs: EST734946
            Contact: Robin Buell
            The Institute for Genomic Research
            9712 Medical Center Dr, Rockville, MD 20850, USA
            Email: potato-array@tigr.org
            Clones can be requested from the University of Arizona Genomics
            Institute via http://genome.arizona.edu/orders/ .
            Seq primer: ATT TAG GTG ACA CTA TAG.
FEATURES             Location/Qualifiers
     source          1. .958
                     /organism="Solanum tuberosum"
                     /mol_type="mRNA"
                     /cultivar="Kennebec"
                     /db_xref="taxon:4113"
                     /clone="POCBM46"
                     /tissue_type="callus"
                     /lab_host="DH10B-TonA"
                     /clone_lib="potato callus cDNA library, normalized and
                     full-length"
                     /note="Vector: pCMVSport6.1; Site_1: EcoRI; Site_2: NotI;
                     supplier: RNA was isolated from Solanum tuberosum var.
                     Kennebec callus tissue grown on solid media."
ORIGIN

Alignment Scores:
Pred. No.:              4.54e-125       Length:         958
Score:                  1309.00         Matches:        232
Percent Similarity:     83.97%          Conservative:   30
Best Local Similarity:  74.36%          Mismatches:     50
Query Match:            31.56%          Indels:         0
DB:                     7               Gaps:           0

US-09-419-679-6 (1-757) x CK251308 (1-958)

Qy          1 MetTrpLysLeuLysPheAlaGluGlyGlyAsnProTrpLeuArgThrLeuAsnAsnHis 20

Db         22 ATGTGGAAGTTGAAAGTAGCAGAAGGAGGGAGTCCATGGCTGCGGACAACGAACAATCAT 81

Qy         21 ValGlyArgGlnValTrpGluPheAspProLysLeuGlySerProGlnAspLeuLeuGlu 40

Db         82 GTAGGTAGGGAAGTATGGGAGTTTGAACCAGACCTTGGATCTCCAGAAGATCGGGCGGAG 141

Qy         41 IleGluLysAlaArgGlnAsnPheHisAspAsnArgPheThrHisLysHisSerAlaAsp 60

Db        142 ATCGAAAAGTTTCGTGAACACTTCACCAAACATCGTTTGGAGCAAAAACACAGTGCTGAT 201

Qy         61 LeuLeuMetArgMetGlnPheAlaArgGluAsnProThrArgGluValLeuProLysVal 80

Db        202 CTCCTTATGCGTTACCAGTTGTCAAAGGAGAACCAGGGAATTAGTATCTTGCCACAAGTC 261

Qy         81 GlyValLysAspIleGluAspValThrGlnGluIleValThrLysThrLeuArgArgAla 100

Db        262 AAAATACATGACAATGAAGTTATTACAGAAGATAGTGTAGCTACCACGCTGAGAAGAGCG 321

Qy        101 ValSerPheHisSerThrLeuGlnCysHisAspGlyHisTrpProGlyAspTyrGlyGly 120

Db        322 CTAAGTTTTTATTCCACTCTACAAAATCATGACGGTCACTGGGCTGGAGATTATGGAGGC 381

Qy        121 ProMetPheLeuMetProGlyLeuValIleThrLeuSerIleThrGlyAlaLeuAsnThr 140

Db        382 CCGATGTTTCTAATGCCTGGCATGATTATTGCTCTATCTATTACGGGGGCACTCAATGCA 441

Qy        141 ValLeuThrGluGluHisArgLysGluIleCysArgTyrLeuTyrAsnHisGlnAsnLys 160

Db        442 GTATTCTCAAGTGAACATAAACGTGAGATGATTCGATATCTTTATAACCATCAGAACAGT 501

Qy        161 AspGlyGlyTrpGlyLeuHisIleGluGlyProSerThrMetPheGlySerValLeuSer 180

Db        502 GATGGTGGGTGGGGTTTGCATATTGAAAGCCACAGTACCATGTTTGGTTCTGCTTTGAGC 561

Qy        181 TyrIleThrLeuArgLeuLeuGlyGluGlyProAsnAspGlyGlnGlyGluMetGluLys 200

Db        562 TATGTTACTCTGAGATTGCTTGGGGAAGGAGCTAATGATGGAGATGGGGCAATGGAAAAG 621

Qy        201 AlaArgAspTrpIleLeuGlyHisGlyGlyAlaThrTyrIleThrSerTrpGlyLysMet 220

Db        622 GGCCGTAAATGGATTCTAGATCATGGTAGTGCCACGGCCATTACGTCATGGGGGAAAATG 681

Qy        221 TrpLeuSerValLeuGlyValTyrGluTrpSerGlyAsnAsnProLeuProProGluIle 240

Db        682 TGGCTCTCAGTGCTTGGAGTATTTGACTGGTCTGGGAACAATCCACTCCCTCCTGAAATA 741

Qy        241 TrpLeuLeuProTyrMetLeuProPheHisProGlyArgMetTrpCysHisCysArgMet 260

Db        742 TGGCTTCTTCCCTATATCCTTCCAATCCATCCTGGAAGAATGTGGTGCCATTGCCGGATG 801

Qy        261 ValTyrLeuProMetSerTyrLeuTyrGlyLysArgPheValGlyProIleSerProThr 280

Db        802 GTTTATCTTCCTATGTGCTACCTTTATGGTAAACGCTTTGTTGGTCCTATCACGCCAACA 861

Qy        281 ValLeuSerLeuArgLysGluLeuTyrThrValProTyrHisAspIleAspTrpAspGln 300

Db        862 GTTTTATCTTTGAGAAAGGAGCTCTTTACGGTCCCGTATCATGGAGTAGATTTGGGATAA 921

Qy        301 AlaArgAsnLeuCysAlaLysGluAspLeuTyrTyr 312

Db        922 GCTCGCAATGAGTGTGCAAAGGAAGACCTTTACTAC 957

RESULT 10
CD440985
LOCUS       CD440985                 850 bp    mRNA    linear   EST 03-JUN-2003
DEFINITION  EL01N0563E07.b Endosperm_5 Zea mays cDNA, mRNA sequence.
ACCESSION   CD440985
VERSION     CD440985.1  GI:31356628
KEYWORDS    EST.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 850)
```

```
Db        302 ATGAAGGGGTACAATGGAAGTCAATCTTGGGATACATCATTTGCAGTTCAAGCAATCATT 361

Qy        427 AlaSerAsnLeuIleGluGluPheGlyProThrIleArgLysAlaHisThrTyrIleLys 446
              |||:::       ||||||:::|||   |||:::||||||||||||:::::::::
Db        362 GCGACAGGGTTTGGCGAAGAATATGGATCAACATTGAGAAAAGCACATTCATTCTTATCA 421

Qy        447 AsnSerGlnValLeuGluAspCysProGlyAspLeuAsnLysTrpTyrArgHisIleSer 466
              |||:::||||||||:::::::||||||||||||||:::   ||||||||||||||||||
Db        422 AATACACAAGTGTTAGATAATTGCCCGGGTGATCTAGATTTCTGGTATCGTCATATTTCA 481

Qy        467 LysGlyAlaTrpProPheSerThrGlyAspHisGlyTrpProIleSerAspCysThrAla 486
              ||||||||||||||||||||||||   ||||||||||||||||||||||||||||||||
Db        482 AAAGGTGCTTGGCCCTTTTCAACTGCAGATCATGGGTGGCCCATTTCAGATTGTACTGCA 541

Qy        487 GluGlyLeuLysAlaValLeuLeuLeuSerLysIleAlaProGluIleValGlyGluPro 506
              ||||||||||||||||||||||||||||:::     ||||||||||||   |||
Db        542 GAGGGACTTAAGGCTGTTCTTCTATTATCAAAGCTACCATCAGAGATAGTCGGTAATCCA 601

Qy        507 IleAspValLysArgLeuTyrAspSerValAsnValIleLeuSerLeuGlnAsnGluAsp 526
              ::::::   ||||||||||||||:::||||||||:::|||||||||||||||
Db        602 TTGGAGGCGAAGCGCTTGTATGATGCTGTAAATGTTTTACTTTCTTTACAGAATAGTGGT 661

Qy        527 GlyGlyPheAlaThrTyrGluLeuLysArgSerTyrAsnTrpLeuGluIleIleAsnPro 546
              ||||||||||||||||||||||||   |||||||||   |||||||||:::|||||||||
Db        662 GGTGGCTTTGCAACATATGAACTCTCAAGGTCATATCCATGGTTGGAGCTAATCAACCCT 721

Qy        547 AlaGluThrPheGlyAspIleValIleAspTyrProTyrValGluCysThrSerAlaAla 566
              ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db        722 GCTGAGACTTTTGGAGATATTGTTATTGATTACCCTTATGTAGAGTGTTCCTCAGCTGCA 781

Qy        567 IleGlnAlaLeuAlaSerPheArgLysLeuTyrProGlyHisArgArgGluGluIleGln 586
              ||||||||||||   :::|||:::||||||||||||||||||||:::||||||::::::
Db        782 ATTCAAGCTTTAACAGCATTTAAGAAATTATACCCTGGTCATCGTAAAGAAGAAGTTGAA 841

Qy        587 HisCysIleAspLysAlaThrThrPheIle-GluLysIleGlnAlaSerAspGlySer-T 606
                  ||||||  ||||||       |||||| ::::::         :::|||||||| |
Db        842 CGCTGTATTGCTAAAGCTGCCGCCTTCATTGGATCAATTTTCAGAAACAGATGGCTCCCT 901

Qy        606 rpTyrGlySerTrpGlyValCysPheThrTyrGlyAlaTrp 619
              |||||||||||||   |||||||||||||||||   |||
Db        902 GGTATGGCTCTTGGGCAGTTTGCTTCACCTATGGCACTTGG 942
```

RESULT 8
BG588039

```
LOCUS       BG588039                859 bp    mRNA    linear   EST 11-APR-2001
DEFINITION  EST489814 MHAM Medicago truncatula/Glomus versiforme mixed EST
            library cDNA clone pMHAM-51P20 5' end, mRNA sequence.
ACCESSION   BG588039
VERSION     BG588039.1  GI:13603103
KEYWORDS    EST.
SOURCE      Medicago truncatula/Glomus versiforme mixed EST library
  ORGANISM  Medicago truncatula/Glomus versiforme mixed EST library
            Eukaryota; mixed EST libraries.
REFERENCE   1  (bases 1 to 859)
  AUTHORS   Harrison,M.J., Liu,J., Town,C.D., Van Aken,S., Utterback,T., Cho,J.
            and Fraser,C.M.
  TITLE     ESTs from roots of Medicago truncatula after colonization with
            Glomus versiforme, 2001
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Harrison M.J.
            Plant Biology Division
            The Samuel Roberts Noble Foundation
            2510 Sam Noble Parkway, Ardmore, OK 73401
            Tel: 580-223-5810
            Fax: 580-221-7380
            Email: mjharrison@noble.org
            Noble EST name: N387112e TIGR sequence name: MTDDN94TK More
            information is available at: http://www.medicago.org
            Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
FEATURES             Location/Qualifiers
     source          1. .859
                     /organism="Medicago truncatula/Glomus versiforme mixed EST
                     library"
                     /mol_type="mRNA"
                     /cultivar="Medicago truncatula genotype A17"
                     /db_xref="taxon:119092"
                     /clone="pMHAM-51P20"
                     /tissue_type="roots colonized with Glomus versiforme"
                     /dev_stage="Roots harvested at 10, 17, 22, 31 and 38 days
                     post-inoculation with Glomus versiforme. The library was
                     made from a mixture of RNA from each of these stages."
                     /lab_host="E. coli strain XLOLR"
                     /clone_lib="MHAM"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; cDNA was prepared from polyA+ enriched RNA from
                     roots harvested at 10, 17, 22, 31 and 38 days
                     post-inoculation with Glomus versiforme. The cDNA was
                     directionally ligated into the Unizap XR vector from
                     Stratagene and packaged using Gigapack III Gold packaging
                     extracts. Plasmids containing cDNA inserts were excised
                     from the recombinant lambda-Zap phage using Ex-assist
                     helper phage and propagated in XLOLR cells."
ORIGIN
```

Alignment Scores:

Pred. No.:	6.41e-132	Length:	859
Score:	1374.00	Matches:	253
Percent Similarity:	93.73%	Conservative:	16
Best Local Similarity:	88.15%	Mismatches:	17
Query Match:	33.12%	Indels:	1
DB:	4	Gaps:	0

US-09-419-679-6 (1-757) x BG588039 (1-859)

```
Qy        408 GlnGlyTyrAsnGlySerGlnLeuTrpAspThrAlaPheAlaValGlnAlaIleIleAla 427
              ||||||||||||||||||||||||||||||||||:::|||      ||||||||||||:::
Db          1 CAGGGCTACAATGGCAGTCAACTATGGGATACTTCATTTATTGCCCAAGCAATTATTTCA 60

Qy        428 SerAsnLeuIleGluGluPheGlyProThrIleArgLysAlaHisThrTyrIleLysAsn 447
              :::|||||||||||||||||||||||||||:::|||||||||||||||:::|||||||||
Db         61 ACTAACCTAATTGAAGAATTTGGTCCAACTCTAAGAAAGGCTCATACATTCATTAAGAAT 120

Qy        448 SerGlnValLeuGluAspCysProGlyAspLeuAsnLysTrpTyrArgHisIleSerLys 467
              |||||||||  ||||||||||||||||||||:::|||||||||||||||||||||||||
Db        121 TCACAGGTTTCAGAAGATTGTCCTGGTGATCTTAGTAAATGGTACCGTCACATTTCAAAA 180

Qy        468 GlyAlaTrpProPheSerThrGlyAspHisGlyTrpProIleSerAspCysThrAlaGlu 487
              ||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db        181 GGTGCTTGGCCCTTTTCAACTGCAGATCATGGATGGCCAATATCTGACTGCACAGCTGAA 240

Qy        488 GlyLeuLysAlaValLeuLeuLeuSerLysIleAlaProGluIleValGlyGluProIle 507
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        241 GGACTGAAAGCTGTCCTATTGCTATCTAAAATTGCACCTGAGATTGTTGGTGAACCATTG 300

Qy        508 AspValLysArgLeuTyrAspSerValAsnValIleLeuSerLeuGlnAsnGluAspGly 527
              |||  |||||||||||||:::|||||||||||:::|||||||||||||||||||||||
Db        301 GATGCAAAGAGATTATATGATGCTGTAAATGTCATTATCTCATTACAGAATGAAGATGGT 360

Qy        528 GlyPheAlaThrTyrGluLeuLysArgSerTyrAsnTrpLeuGluIleIleAsnProAla 547
              |||   ||||||||||||||   |||||||||:::||||||||||||   |||||||||
Db        361 GGCCTTGCAACATATGAGCTCACACGATCTTATAGCTGGTTGGAGATATGCAATCCTGCT 420

Qy        548 GluThrPheGlyAspIleValIleAspTyrProTyrValGluCysThrSerAlaAlaIle 567
              |||||||||||||||||||||||||||||||||||   ||||||||||||||||||||||
Db        421 GAAACTTTTGGCGACATTGTTATTGATTATACTTATGTGGAATGTACCTCAGCAGCAATT 480

Qy        568 GlnAlaLeuAlaSerPheArgLysLeuTyrProGlyHisArgArgGluGluIleGlnHis 587
              |||||||||::::::|||||||||||||||||||||||||||||||||||||||||||
Db        481 CAGGCTTTGTCAACATTTAGGAAACTATATCCTGGGCATCGTCGGGAAGAAATACAACGC 540

Qy        588 CysIleAspLysAlaThrThrPheIleGluLysIleGlnAlaSerAspGlySerTrpTyr 607
              ||||||:::|||||||      ||||||||||||||||||||||||||||||||||||
Db        541 TGTATAGAAAAAGCTGCTGCCTTTATTGAAAAAATACAAGCTTCAGATGGTTCATGGTAT 600

Qy        608 GlySerTrpGlyValCysPheThrTyrGlyAlaTrpPheGlyValLysGlyLeuIleAla 627
              ||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||
```

```
Db       1192 TATGAAGCTTATTACCGCCATAGATCAAAAGGCTCATGGACGCTTTCAACAGCAGATAAT 1251

Qy        478 GlyTrpProIleSerAspCysThrAlaGluGlyLeuLysAlaValLeuLeuLeuSerLys 497

Db       1252 GGATGGTGTGTATCTGATTGTACAGCTGAAGCACTTAAGGCATTACTTATGCTATCAAAG 1311

Qy        498 IleAlaProGluIleValGlyGluProIleAspValLysArgLeuTyrAspSerValAsn 517

Db       1312 ATCTCTCAAGATCTAGTGGGAGATCCGATAGACGGAGAAAGGTTGTATGATGCAGTGGAT 1371

Qy        518 ValIleLeuSerLeu---------------------------------------------- 522

Db       1372 GGCATGCTCTCTTTCATGACATACCAGTTGTGGGGGCTAATACAGATTTTGGAAGCTGAC 1431

Qy        523 GlnAsnGluAspGlyGlyPheAlaThrTyrGluLeuLysArgSerTyrAsnTrpLeuGlu 542

Db       1432 ATGAATGAAGATGGCACCTTTTCTACATATGAGTGCAAGAGAAGTACACCTTGGTTAGAG 1491

Qy        543 IleIleAsnProAlaGluThrPheGlyAspIleValIleAspTyrProTyrValGluCys 562

Db       1492 GTTCTCAACCCGTCAGAGAGTTTTCTGAACATTGTAGTTGATTAT---------------- 1536

Qy        563 ThrSerAlaAlaIleGlnAlaLeuAlaSerPheArgLysLeuTyrProGlyHisArgArg 582

Db       1536 ------------------------------------------------------------ 1536

Qy        583 GluGluIleGlnHisCysIleAspLysAlaThrThrPheIleGluLysIleGlnAlaSer 602

Db       1536 ------------------------------------------------------------ 1536

Qy        603 AspGlySerTrpTyrGlySerTrpGlyValCysPheThrTyrGlyAlaTrpPheGlyVal 622

Db       1537 ---------CTGTTTGGTACTTGGGGTATATGTTTCACTTATGGGACATTCTTTGCTGTA 1587

Qy        623 LysGlyLeuIleAlaAlaGlyArgSerPheSerAsnCysSerSerIleArgLysAlaCys 642

Db       1588 AAAGGATTAGCTGCTACTGGAAGAACATACGAAAATAGCTCTTCCATTAGAAAAGCATGC 1647

Qy        643 GluPheLeuLeuSerLysGlnLeuProSerGlyGlyTrpGlyGluSerTyrLeuSerCys 662

Db       1648 AACTTTATTTTGTCAAAGCAACTTAGTACAGGTGGATGGGGAGAGACTTATCTTTCTAGT 1707

Qy        663 GlnAsnLysValTyrSerAsnLeuGluGlyAsnArgSerHisValValAsnThrGlyTrp 682

Db       1708 GAAACTGAGGTTTAT------GTGGAGGCTAGTAGACCTCATGCAGTGAACACTGCTTGG 1761

Qy        683 AlaMetLeuAlaLeuIleAspAlaGlyGlnAlaLysArgAspSerGlnProLeuHisArg 702

Db       1762 GCAATGTTAGCTTTAATATATGCTGGGCAGGTTGAAATAGATCCAACACCACTATATCAT 1821

Qy        703 AlaAlaAlaTyrLeuIleAsnSerGlnLeuGluAspGlyAspProProGlnGln 720

Db       1822 GCAGCACAAGAATTGATCAATATGCAACTAGAAACAGGCGAGTTTCCACAACAA 1875
```

RESULT 7
CK265941
LOCUS CK265941 943 bp mRNA linear EST 03-AUG-2004
DEFINITION EST712019 potato abiotic stress cDNA library Solanum tuberosum cDNA clone POAC172 5' end, mRNA sequence.
ACCESSION CK265941
VERSION CK265941.1 GI:39822919
KEYWORDS EST.
SOURCE Solanum tuberosum (potato)
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; lamiids; Solanales; Solanaceae; Solanum.
REFERENCE 1 (bases 1 to 943)
AUTHORS Buell,C.R., Hart,A., Zismann,V., Karamycheva,S.A. and Baker,B.
TITLE Generation of ESTs from abiotic stressed potato tissue
JOURNAL Unpublished (2003)
COMMENT Other_ESTs: EST712020
Contact: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr, Rockville, MD 20850, USA
Email: potato-array@tigr.org
Clones can be requested from the University of Arizona Genomics Institute via http://genome.arizona.edu/orders/ .
Seq primer: ATT TAG GTG ACA CTA TAG.

FEATURES Location/Qualifiers
source 1. .943
/organism="Solanum tuberosum"
/mol_type="mRNA"
/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="POAC172"
/tissue_type="abiotic stress treated leaf and root tissue"
/lab_host="DH10B-TonA"
/clone_lib="potato abiotic stress cDNA library"
/note="Vector: pCMVSport6.1; Site_1: EcoRI; Site_2: NotI; supplier: Solanum tuberosum var. Kennebec plants were grown from cuttings on a 16hr light/8 hr dark cycle at 25 C for 3-4 weeks. Abiotic stress conditions were applied to four separate sets of plants. Set 1 involved saturation of the soil with 150 mM NaCl and tissues were harvested at following application of the salt stress (leaves: 2hr, 6hr, 12hr, 1d, 2d, and 4d; roots:2hr, 6hr, 12hr, and 2d). Set 2 were grown under the standard conditions and then were water stressed by withdrawal of further watering applications. Drought stressed plants were harvested after cessation of watering (leaves: 3d, 5d, and 7d; roots:3d and 5d). Set 3 were grown under the standard conditions and then were cold stressed by placement at 4 C. Cold stressed leaves were harvested at 2 hr, 6 hr, 12 hr, 1 d, and 4d and roots were harvested at 2 hr, 6 hr, 12 hr, 1 d, 2d. Set 4 were grown under the standard conditions and then were heat stressed by placement at 35 C. Heat stressed leaves were harvested at 2 hr, 6 hr, 12 hr, 1 d, 2d and 4d and heat-stressed roots were harvested at 6 hr, 12 hr, 1 d, and 4d. RNA was isolated from all tissues and equal RNA from each tissue and stress was pooled to construct the cDNA library. RNA sample."

ORIGIN

Alignment Scores:

Pred. No.:	2.72e-135	Length:	943
Score:	1407.00	Matches:	244
Percent Similarity:	90.16%	Conservative:	40
Best Local Similarity:	77.46%	Mismatches:	29
Query Match:	33.92%	Indels:	2
DB:	7	Gaps:	0

US-09-419-679-6 (1-757) x CK265941 (1-943)

```
Qy        307 LysGluAspLeuTyrTyrProHisProLeuValGlnAspIleLeuTrpAlaSerLeuHis 326

Db          2 AAGGAAGACCTTTACTACCCACACCCTCTAGTACAAGATATCATGTGGGCAACTCTTCAC 61

Qy        327 LysPheLeuGluProIleLeuMetHisTrpProGlyLysArgLeuArgGluLysAlaIle 346

Db         62 AAGGTTGTAGAACCTATTTTGATGCATTGGCCGGGGAAAAGATTGAGGGAAAAATCTCTG 121

Qy        347 IleSerAlaLeuGluHisIleHisTyrGluAspGluAsnThrArgTyrIleCysIleGly 366

Db        122 CGCACTGTGATGGAGCACATTCATTATGAAGATGAAAACACTCGTTACGTATGCATAGGA 181

Qy        367 ProValAsnLysValLeuAsnMetLeuCysCysTrpValGluAspProAsnSerGluAla 386

Db        182 CCTGTAAACAAGATTATAAATATGCTTTGCTGCTGGGTTGAAGATCCTAATTCAGAAGCT 241

Qy        387 PheLysLeuHisLeuProArgIleTyrAspTyrLeuTrpIleAlaGluAspGlyMetLys 406

Db        242 TTCAAGTTGCATCTTCCAAGAATTCATGATTATATGTGGGTTGCTGAAGATGGCATGAAA 301

Qy        407 MetGlnGlyTyrAsnGlySerGlnLeuTrpAspThrAlaPheAlaValGlnAlaIleIle 426
```

```
Qy        746 LeuGlyGluTyrGlnSer 751
              ||||||   :::   :::
Db       2193 CTGGGCCGCTTCTCCAAC 2210

RESULT 6
CL964361
LOCUS       CL964361               1887 bp    DNA     linear   GSS 21-SEP-2004
DEFINITION  OsIFCC010736 Oryza sativa Express Library Oryza sativa (indica
            cultivar-group) genomic, genomic survey sequence.
ACCESSION   CL964361
VERSION     CL964361.1  GI:52383428
KEYWORDS    GSS.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 1887)
  AUTHORS   Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
            Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
            Wong,G.K.S., Deng,X.W. and Wang,J.
  TITLE     An analysis of transcriptional regulation of the rice genome and
            its comparison to Arabidopsis
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Chen Chen
            Department of Bioinformatic
            Beijing Institute of Genomics
            Chinese Academy of Sciences, Beijing 101300, China
            Tel: 86-10-80481559
            Fax: 86-10-80488676
            Email: chenchen@genomics.org.cn
            Rice genomic sequence.
            Class: exon-trapped.
FEATURES             Location/Qualifiers
     source          1. .1887
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:39946"
                     /clone_lib="Oryza sativa Express Library"
                     /note="Oryza sativa exon trapped genomic sequences "
ORIGIN

Alignment Scores:
Pred. No.:              2.53e-160       Length:         1887
Score:                  1652.00         Matches:        314
Percent Similarity:     56.64%          Conservative:   104
Best Local Similarity:  42.55%          Mismatches:     122
Query Match:            39.83%          Indels:         198
DB:                     9               Gaps:           7

US-09-419-679-6 (1-757) x CL964361 (1-1887)

Qy         67 PheAlaArgGluAsnProThrArgGluValLeuProLysValGlyValLysAspIleGlu 86
              :::|||:::::::|||      :::      :::|||         :::   |||
Db          4 TATGCAAAACAAAATCGCCTTCAGGTGGATGTTCCAGCCAGTAAACTTGTCGATAGTACT 63

Qy         87 AspValThrGlnGluIleValThrLysThrLeuArgArgAlaValSerPheHisSerThr 106
                 ||||||:::::::|||:::      :::|||||||||||||:::::::   ||||||:::
Db         64 CAAGTCACAGAACAGATTATATTGGCATCACTGAGGCGGGCTCTTACTCAACATTCTTCT 123

Qy        107 LeuGlnCysHisAspGlyHisTrpProGlyAspTyrGlyGlyProMetPheLeuMetPro 126
              ||||||   |||||||||||||||||||||||:::   |||   ||||||:::||||||
Db        124 CTGCAAGCACATGATGGACATTGGCCTGGCGATTTCAGTGGAATTATGTTCATTATGCCT 183

Qy        127 GlyLeuValIleThrLeuSerIleThrGlyAlaLeuAsnThrValLeuThrGluGluHis 146
                 |||      ::::::
Db        184 ATGTTGGAGAGGAGTATTGAT---------------------------------------- 204

Qy        147 ArgLysGluIleCysArgTyrLeuTyrAsnHisGlnAsnLysAspGlyGlyTrpGlyLeu 166

Db        204 ---------------------------------------------------------- 204

Qy        167 HisIleGluGlyProSerThrMetPheGlySerValLeuSerTyrIleThrLeuArgLeu 186

Db        204 ---------------------------------------------------------- 204

Qy        187 LeuGlyGluGlyProAsnAspGlyGlnGlyGluMetGluLysAlaArgAspTrpIleLeu 206
                                                  :::   |||   |||   ||||||:::
Db        205 ----------------------------GCACTAACCAAAGGGCGTGCTTGGATTATA 234

Qy        207 GlyHisGlyGlyAlaThrTyrIleThrSerTrpGlyLysMetTrpLeuSer--------- 223
                 |||   |||      :::      |||||||||:::|||||||||
Db        235 TCCCGTGGGAGTGCAGCTGCAGTACCACAGTGGGGAAAGATATGGTTGTCGTTTATCAAG 294

Qy        224 ---------------------------------------------ValLeuGlyValTyr 228
                                                          |||:::|||:::|||
Db        295 TTGTATGATGAAAGGGTTAATGCAAGCTATTCATATATTCTGAAGGTGATTGGTTTGTAT 354

Qy        229 GluTrpSerGlyAsnAsnProLeuProProGluIleTrpLeuLeuProTyrMetLeuPro 248
              :::|||||||||||||||   :::   ||||||:::||||||:::|||:::   ||||||
Db        355 GATTGGTCTGGAAATAATGCCATCATTCCTGAATTGTGGCTTGTACCACATTTTCTTCCA 414

Qy        249 PheHisProGlyArgMetTrpCysHisCysArgMetValTyrLeuProMetSerTyrLeu 268
              ||||||   |||   ||||||   ||||||:::||||||:::||||||:::||||||
Db        415 ATTCACCCAGAGCGATTCTGGTGCTTCTGTCGTTTGGTTTATATGCCAATGGCTTATCTT 474

Qy        269 TyrGlyLysArgPheValGlyProIleSerProThrValLeuSerLeuArgLysGluLeu 288
              |||||||||:::|||||||||   |||:::||||||:::|||   :::|||:::||||||
Db        475 TATGGTAAGAAGTTTGTTGGGGCAATTACACCAACGATATTGGAAATAAGAGAGGAGCTC 534

Qy        289 TyrThrValProTyrHisAspIleAspTrpAspGlnAlaArgAsnLeuCysAlaLysGlu 308
              |||:::|||||||||:::::::|||:::|||      |||||||||   ||||||||||
Db        535 TACAGTGTACCATATAATGAGATTAACTGGAAAAATGCTCGGAACAATTGTGCCAAGGAA 594

Qy        309 AspLeuTyrTyrProHisProLeuValGlnAspIleLeuTrpAlaSerLeuHisLysPhe 328
              ||||||   ||||||         ||||||:::::::::|||      |||:::|||
Db        595 GACCTTCGTTATCCGCGATCATTTGTACAAAATGTTATTTGGACTGGTCTTAACAAAGTT 654

Qy        329 LeuGluProIleLeuMetHisTrpProGlyLysArgLeuArgGluLysAlaIleIleSer 348
              :::||||||||||||:::   ||||||         ||||||      |||:::   :::
Db        655 GTGGAACCAATATTGATCCTTTGGCCATTTAATACCTTAAGACACGCTGCACTGAATAAT 714

Qy        349 AlaLeuGluHisIleHisTyrGluAspGluAsnThrArgTyrIleCysIleGlyProVal 368
              |||:::||||||   ||||||||||||:::|||:::||||||   |||   |||:::
Db        715 CTCTTGAAGCATATTCGCTATGAAGATGAATCTACAAAATATATTGGCATATGTCCAATC 774

Qy        369 AsnLysValLeuAsnMetLeuCysCysTrpValGluAspProAsnSerGluAlaPheLys 388
              ||||||   |||:::|||:::|||||||||:::::::::||||||||||:::|||||||||
Db        775 AACAAGGCACTCGACATGATTTGTTGTTGGATAGATAATCCAAATTCTGATGCATTCAAG 834

Qy        389 LeuHisLeuProArgIleTyrAspTyrLeuTrpIleAlaGluAspGlyMetLysMetGln 408
              ||||||||||||||||||||||||||||||:::||||||||||||||||||   |||
Db        835 TTGCATCTTCCAAGAATATACGATTATTTATGGCTTGCAGAAGATGGCATGAAGGCACAG 894

Qy        409 Gly---------------------------------------------------------- 409

Db        895 CATATACTTCCTTCACTTCAAAATTTCCCTGGTTGTGGTCACAATGGCGGAGATGGCTTA 954

Qy        409 ---------------------------------------------------------- 409

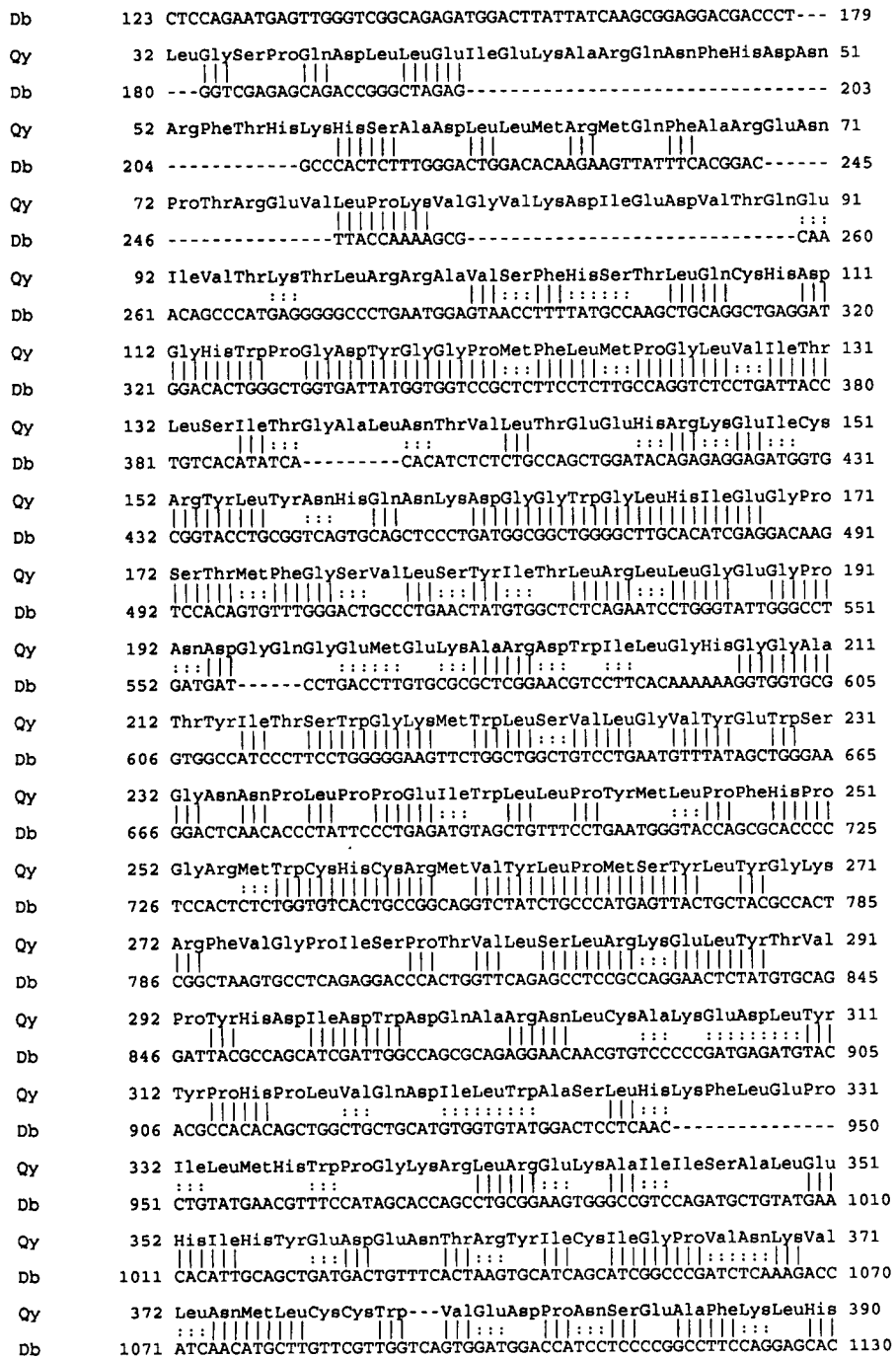
Db        955 GGCATTGGAGCACACTCGCGGTGGTCATGGCGTGGGAGGCAGCAAGCTCATGGAGCAGGC 1014

Qy        410 ---------------------------------TyrAsnGlySerGlnLeuTrpAsp 417
                                                  |||   |||   |||   |||:::
Db       1015 GGTGCAGGTAGCTTCAAGATCAAGTTGGAGATGGTGTATGGTGGTTGTCAAAGCTGGGAA 1074

Qy        418 ThrAlaPheAlaValGlnAlaIleIleAlaSerAsnLeuIleGluGluPheGlyProThr 437
              |||||||||   |||||||||      :::::::::|||:::   ||||||      |||
Db       1075 ACAGCATTTATTGTTCAAGCTTATTGTTCAACAGACCTTGTTAATGAGTTTAGTCAAACA 1134

Qy        438 IleArgLysAlaHisThrTyrIleLysAsnSerGlnValLeuGluAspCysProGlyAsp 457
              :::   |||||||||   :::||||||   |||||||||||||||:::   |||   |||
Db       1135 CTTACAAAAGCACACGAGTTCATAAAAAAGTCACAGGTCCTTGAAAACCATCCA---GAC 1191

Qy        458 LeuAsnLysTrpTyrArgHisIleSerLysGlyAlaTrpProPheSerThrGlyAspHis 477
                   :::|||||||||   |||||||||:::|||      ||||||   |||:::
```


Db 123 CTCCAGAATGAGTTGGGTCGGCAGAGATGGACTTATTATCAAGCGGAGGACGACCCT--- 179
Qy 32 LeuGlySerProGlnAspLeuLeuGluIleGluLysAlaArgGlnAsnPheHisAspAsn 51
Db 180 ---GGTCGAGAGCAGACCGGGCTAGAG---------- 203
Qy 52 ArgPheThrHisLysHisSerAlaAspLeuLeuMetArgMetGlnPheAlaArgGluAsn 71
Db 204 ----------GCCCACTCTTTGGGACTGGACACAAGAAGTTATTTCACGGAC------ 245
Qy 72 ProThrArgGluValLeuProLysValGlyValLysAspIleGluAspValThrGlnGlu 91
Db 246 --------------TTACCAAAAGCG----------CAA 260
Qy 92 IleValThrLysThrLeuArgArgAlaValSerPheHisSerThrLeuGlnCysHisAsp 111
Db 261 ACAGCCCATGAGGGGGCCCTGAATGGAGTAACCTTTTATGCCAAGCTGCAGGCTGAGGAT 320
Qy 112 GlyHisTrpProGlyAspTyrGlyGlyProMetPheLeuMetProGlyLeuValIleThr 131
Db 321 GGACACTGGGCTGGTGATTATGGTGGTCCGCTCTTCCTCTTGCCAGGTCTCCTGATTACC 380
Qy 132 LeuSerIleThrGlyAlaLeuAsnThrValLeuThrGluGluHisArgLysGluIleCys 151
Db 381 TGTCACATATCA---------CACATCTCTCTGCCAGCTGGATACAGAGAGGAGATGGTG 431
Qy 152 ArgTyrLeuTyrAsnHisGlnAsnLysAspGlyGlyTrpGlyLeuHisIleGluGlyPro 171
Db 432 CGGTACCTGCGGTCAGTGCAGCTCCCTGATGGCGGCTGGGGCTTGCACATCGAGGACAAG 491
Qy 172 SerThrMetPheGlySerValLeuSerTyrIleThrLeuArgLeuLeuGlyGluGlyPro 191
Db 492 TCCACAGTGTTTGGGACTGCCCTGAACTATGTGGCTCTCAGAATCCTGGGTATTGGGCCT 551
Qy 192 AsnAspGlyGlnGlyGluMetGluLysAlaArgAspTrpIleLeuGlyHisGlyGlyAla 211
Db 552 GATGAT------CCTGACCTTGTGCGCGCTCGGAACGTCCTTCACAAAAAAGGTGGTGCG 605
Qy 212 ThrTyrIleThrSerTrpGlyLysMetTrpLeuSerValLeuGlyValTyrGluTrpSer 231
Db 606 GTGGCCATCCCTTCCTGGGGGAAGTTCTGGCTGGCTGTCCTGAATGTTTATAGCTGGGAA 665
Qy 232 GlyAsnAsnProLeuProProGluIleTrpLeuLeuProTyrMetLeuProPheHisPro 251
Db 666 GGACTCAACACCCTATTCCCTGAGATGTAGCTGTTTCCTGAATGGGTACCAGCGCACCCC 725
Qy 252 GlyArgMetTrpCysHisCysArgMetValTyrLeuProMetSerTyrLeuTyrGlyLys 271
Db 726 TCCACTCTCTGGTGTCACTGCCGGCAGGTCTATCTGCCCATGAGTTACTGCTACGCCACT 785
Qy 272 ArgPheValGlyProIleSerProThrValLeuSerLeuArgLysGluLeuTyrThrVal 291
Db 786 CGGCTAAGTGCCTCAGAGGACCCACTGGTTCAGAGCCTCCGCCAGGAACTCTATGTGCAG 845
Qy 292 ProTyrHisAspIleAspTrpAspGlnAlaArgAsnLeuCysAlaLysGluAspLeuTyr 311
Db 846 GATTACGCCAGCATCGATTGGCCAGCGCAGAGGAACAACGTGTCCCCCGATGAGATGTAC 905
Qy 312 TyrProHisProLeuValGlnAspIleLeuTrpAlaSerLeuHisLysPheLeuGluPro 331
Db 906 ACGCCACACAGCTGGCTGCTGCATGTGGTGTATGGACTCCTCAAC---------- 950
Qy 332 IleLeuMetHisTrpProGlyLysArgLeuArgGluLysAlaIleIleSerAlaLeuGlu 351
Db 951 CTGTATGAACGTTTCCATAGCACCAGCCTGCGGAAGTGGGCCGTCCAGATGCTGTATGAA 1010
Qy 352 HisIleHisTyrGluAspGluAsnThrArgTyrIleCysIleGlyProValAsnLysVal 371
Db 1011 CACATTGCAGCTGATGACTGTTTCACTAAGTGCATCAGCATCGGCCCGATCTCAAAGACC 1070
Qy 372 LeuAsnMetLeuCysCysTrp---ValGluAspProAsnSerGluAlaPheLysLeuHis 390
Db 1071 ATCAACATGCTTGTTCGTTGGTCAGTGGATGGACCATCCTCCCCGGCCTTCCAGGAGCAC 1130
Qy 391 LeuProArgIleTyrAspTyrLeuTrpIleAlaGluAspGlyMetLysMetGlnGlyTyr 410
Db 1131 GTTTCTCGGATCAAAGATTACCTTTGGCTGGGCCTTGATGGCATGAAAATGCAGGGCACC 1190
Qy 411 AsnGlySerGlnLeuTrpAspThrAlaPheAlaValGlnAlaIleIleAlaSerAsnLeu 430
Db 1191 AATGGATCACAGATCTGGGACACTTCATTTGCTATCCAAGCACTGTTAGAGGCAGGTGCA 1250
Qy 431 IleGlu------GluPheGlyProThrIleArgLysAlaHisThrTyrIleLysAsnSer 448
Db 1251 CACCACAGACCTGAGTTTCTGCCCTGCCTGCAGAAGGCTCACGAGTTCCTGCGGCTTTCA 1310
Qy 449 GlnValLeuGluAspCysProGlyAspLeuAsnLysTrpTyrArgHisIleSerLysGly 468
Db 1311 CAGGTTCCAGAAAACTGTCCT---GACTATCAGAAGTACTATCGCCATATGCGCAAGGGT 1367
Qy 469 AlaTrpProPheSerThrGlyAspHisGlyTrpProIleSerAspCysThrAlaGluGly 488
Db 1368 GGTTTCTCCTTCAGCACACTGGACTGTGGCTGGATCGTTGCTGACTGCACAGCCGAGGGT 1427
Qy 489 LeuLysAlaValLeuLeuLeuSerLysIleAlaProGluIleValGlyGluProIleAsp 508
Db 1428 TTGAAGGCTGTGCTGCTCCTGCAGAATCAGTGTCCGTCCATCACT---GAGCATATCCCC 1484
Qy 509 ValLysArgLeuTyrAspSerValAsnValIleLeuSerLeuGlnAsnGluAspGlyGly 528
Db 1485 CGAGAGCGGCTGTGCGATGCTGTGGATGTGTTACTGAGCTTGAGGAATGCTGATGGCGGG 1544
Qy 529 PheAlaThrTyrGluLeuLysArgSerTyrAsnTrpLeuGluIleIleAsnProAlaGlu 548
Db 1545 TTCGCTACCTATGAGAAGAAGCGTGGGGGTTACTTGCTGGAGCTGCTGAACCCCTCAGAG 1604
Qy 549 ThrPheGlyAspIleValIleAspTyrProTyrValGluCysThrSerAlaAlaIleGln 568
Db 1605 GTCTTTGGAGACATCATGATTGACTACACGTATGTAGAGTGTACCTCCGCTGTGATGCAG 1664
Qy 569 AlaLeuAlaSerPheArgLysLeuTyrProGlyHisArgArgGluGluIleGlnHisCys 588
Db 1665 GCACTGAAGCATTTCCATGAACACTTCCCAGACTACAGGGCAGCAGAGGTCAGGGAGACC 1724
Qy 589 IleAspLysAlaThrThrPheIleGluLysIleGlnAlaSerAspGlySerTrpTyrGly 608
Db 1725 CTCAATCAAGGCCTGGACTTCTGCCGAAGGAAGCAGAGAGCCGATGGCTCCTGGGAGGGC 1784
Qy 609 SerTrpGlyValCysPheThrTyrGlyAlaTrpPheGlyValLysGlyLeuIleAlaAla 628
Db 1785 TCCTGGGGTGTTTGTTTCACCTACGGCACCTGGTTTGGCTTGGAAGCGTTCGCTTGCATG 1844
Qy 629 GlyArgSerPheSerAsn---------CysSerSerIleArgLysAlaCysGluPheLeu 645
Db 1845 GGGCATACCTACCAAGATGGGGCTGCTTGTGCAGAAGTAGCTCAGGCCTGCAACTTCCTC 1904
Qy 646 LeuSerLysGlnLeuProSerGlyGlyTrpGlyGluSerTyrLeuSerCysGlnAsnLys 665
Db 1905 CTGTCCCAGCAGATGGCGGATGGAGGCTGGGGAGAGGACTTTGAATCATGTGAACAGCGG 1964
Qy 666 ValTyrSerAsnLeuGluGlyAsnArgSerHisValValAsnThrGlyTrpAlaMetLeu 685
Db 1965 CGGTAT------GTGCAGAGTGCTAGGTCCCAGGTCCATAGCACCTGCTGGGCCCTGATG 2018
Qy 686 AlaLeuIleAspAlaGlyGlnAlaLysArgAspSerGlnProLeuHisArgAlaAlaAla 705
Db 2019 GGTTTGATGGCTGTCAGGCATCCCGACATCACTGCTCAG------GAGAGAGGGATCAGA 2072
Qy 706 TyrLeuIleAsnSerGlnLeuGluAspGlyAspPheProGlnGlnGluIleMetGlyVal 725
Db 2073 TGTCTGCTAGGGAGGCAGTTACCCAATGGAGACTGGCCTCAGGAGAACATCTCTGGGGTC 2132
Qy 726 PheAsnLysAsnCysMetIleThrTyrAlaAlaTyrArgAsnIlePheProIleTrpAla 745
Db 2133 TTCAACAAGTCCTGTGCCATCAGCTACACAAGTTACAGAAACATCTTCCCCATCTGGGCC 2192

```
Db       1718 CTCAATCAAGGCCTGGACTTCTGCCGAAGGAAGCAGAGAGCCGATGGCTCCTGGGAGGGC 1777

Qy        609 SerTrpGlyValCysPheThrTyrGlyAlaTrpPheGlyValLysGlyLeuIleAlaAla 628

Db       1778 TCCTGGGGTGTTTGTTTCACCTACGGCACCTGGTTTGGCTTGGAAGCGTTCGCTTGCATG 1837

Qy        629 GlyArgSerPheSerAsn---------CysSerSerIleArgLysAlaCysGluPheLeu 645

Db       1838 GGGCATACCTACCAAGATGGGGCTGCTTGTGCAGAAGTAGCTCAGGCCTGCAACTTCCTC 1897

Qy        646 LeuSerLysGlnLeuProSerGlyGlyTrpGlyGluSerTyrLeuSerCysGlnAsnLys 665

Db       1898 CTGTCCCAGCAGATGGCGGATGGAGGCTGGGGAGAGGACTTTGAATCATGTGAACAGCGG 1957

Qy        666 ValTyrSerAsnLeuGluGlyAsnArgSerHisValValAsnThrGlyTrpAlaMetLeu 685

Db       1958 CGGTAT------GTGCAGAGTGCTAGGTCCCAGGTCCATAGCACCTGCTGGGCCCTGATG 2011

Qy        686 AlaLeuIleAspAlaGlyGlnAlaLysArgAspSerGlnProLeuHisArgAlaAlaAla 705

Db       2012 GGTTTGATGGCTGTCAGGCATCCCGACATCACTGCTCAG------GAGAGAGGGATCAGA 2065

Qy        706 TyrLeuIleAsnSerGlnLeuGluAspGlyAspPheProGlnGlnGluIleMetGlyVal 725

Db       2066 TGTCTGCTAGGGAAGCAGTTACCCAATGGAGACTGGCCTCAGGAGAACATCTCTGGGGTC 2125

Qy        726 PheAsnLysAsnCysMetIleThrTyrAlaAlaTyrArgAsnIleProIleTrpAla 745

Db       2126 TTCAACAAGTCCTGTGCCATCAGCTACACAAGTTACAGAAACATCTTCCCCATCTGGGCC 2185

Qy        746 LeuGlyGluTyrGlnSer 751

Db       2186 CTGGGCCGCTTCTCCAAC 2203
```

RESULT 5
AK031271
LOCUS AK031271 4659 bp mRNA linear HTC 03-APR-2004
DEFINITION Mus musculus 13 days embryo male testis cDNA, RIKEN full-length enriched library, clone:6030400F16 product:OXIDOSQUALENE CYCLASE homolog [Rattus norvegicus], full insert sequence.
ACCESSION AK031271
VERSION AK031271.1 GI:26082233
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6 (bases 1 to 4659)
AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL:http://genome.gsc.riken.jp/
URL:http://fantom.gsc.riken.jp/.
FEATURES Location/Qualifiers
source 1. .4659
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:6030400F16"
/db_xref="taxon:10090"
/clone="6030400F16"
/sex="male"
/tissue_type="testis"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="13 days embryo"
misc_feature 45. .2246
/note="OXIDOSQUALENE CYCLASE homolog [Rattus norvegicus] (SPTR|Q62811, evidence: FASTY, 92.1%ID, 100%length, match=2199)
putative"
ORIGIN

Alignment Scores:

Pred. No.:	7.75e-160	Length:	4659
Score:	1653.50	Matches:	337
Percent Similarity:	61.26%	Conservative:	120
Best Local Similarity:	45.17%	Mismatches:	228
Query Match:	39.86%	Indels:	61
DB:	3	Gaps:	15

US-09-419-679-6 (1-757) x AK031271 (1-4659)

```
Qy         17 LeuAsnAsnHisValGlyArgGlnValTrp---------------GluPheAspProLys 31
```

putative"
/codon_start=1
/protein_id="BAC37102.1"
/db_xref="GI:26346907"
/translation="MTEGTCLRRRGGPYKTEPATDLTRWRLQNELGRQRWTYYQAEDD
PGREQTGLEAHSLGLDTRSYFTDLPKAQTAHEGALNGVTFYAKLQAEDGHWAGDYGGP
LFLLPGLLITCHISHISLPAGYREEMVRYLRSVQLPDGGWGLHIEDKSTVFGTALNYV
ALRILGIGPDDPDLVRARNVLHKKGGAVAIPSWGKFWLAVLNVYSWEGLNTLFPEMWL
FPEWVPAHPSTLWCHCRQVYLPMSYCYATRLSASEDPLVQSLRQELYVQDYASIDWPA
QRNNVSPDEMYTPHSWLLHVAYGLLNLYERFHSTSLRKWAVQMLYEHIAADDCFTKCI
SIGPISKTINMLVRWSVDGPSSPAFQEHVSRIKDYLWLGLDGMKMQGTNGSQIWDTSF
AIQALLEAGAHHRPEFLPCLQKAHEFLRLSQVPENCPDYQKYYRHMRKGGFSFSTLDC
GWIVADCTAEGLKAVLLLQNQCPSITEHIPRERLCDAVDVLLSLRNADGGFATYEKKR
GGYLLELLNPSEVFGDIMIDYTYVECTSAVMQALKHFHEHFPDYRAAEVRETLNQGLD
FCRRKQRADGSWEGSWGVCFTYGTWFGLEAFACMGHTYQDGAACAEVAQACNFLLSQQ
MADGGWGEDFESCEQRRYVQSARSQVHSTCWALMGLMAVRHPDITAQERGIRCLLGKQ
LPNGDWPQENISGVFNKSCAISYTSYRNIFPIWALGRFSNLYPDNTLAGHI"

ORIGIN

Alignment Scores:

Pred. No.:	1.2e-161	Length:	2836
Score:	1667.50	Matches:	338
Percent Similarity:	61.26%	Conservative:	119
Best Local Similarity:	45.31%	Mismatches:	228
Query Match:	40.20%	Indels:	61
DB:	3	Gaps:	15

US-09-419-679-6 (1-757) x AK078023 (1-2836)

```
Qy        17 LeuAsnAsnHisValGlyArgGlnValTrp----------------GluPheAspProLys  31
Db       116 CTCCAGAATGAGTTGGGTCGGCAGAGATGGACTTATTATCAAGCGGAGGACGACCCT---  172

Qy        32 LeuGlySerProGlnAspLeuLeuGluIleGluLysAlaArgGlnAsnPheHisAspAsn  51
Db       173 ---GGTCGAGAGCAGACCGGGCTAGAG------------------------------  196

Qy        52 ArgPheThrHisLysHisSerAlaAspLeuLeuMetArgMetGlnPheAlaArgGluAsn  71
Db       197 ------------GCCCACTCTTTGGGACTGGACACAAGAAGTTATTTCACGGAC------  238

Qy        72 ProThrArgGluValLeuProLysValGlyValLysAspIleGluAspValThrGlnGlu  91
Db       239 ---------------TTACCAAAAGCG----------------------------CAA  253

Qy        92 IleValThrLysThrLeuArgArgAlaValSerPheHisSerThrLeuGlnCysHisAsp  111
Db       254 ACAGCCCATGAGGGGGCCCTGAATGGAGTAACCTTTTATGCCAAGCTGCAGGCTGAGGAT  313

Qy       112 GlyHisTrpProGlyAspTyrGlyGlyProMetPheLeuMetProGlyLeuValIleThr  131
Db       314 GGACACTGGGCTGGTGATTATGGTGGTCCGCTCTTCCTCTTGCCAGGTCTCCTGATTACC  373

Qy       132 LeuSerIleThrGlyAlaLeuAsnThrValLeuThrGluGluHisArgLysGluIleCys  151
Db       374 TGTCACATATCA---------CACATCTCTCTGCCAGCTGGATACAGAGAGGAGATGGTG  424

Qy       152 ArgTyrLeuTyrAsnHisGlnAsnLysAspGlyGlyTrpGlyLeuHisIleGluGlyPro  171
Db       425 CGGTACCTGCGGTCAGTGCAGCTCCCTGATGGCGGCTGGGGCTTGCACATCGAGGACAAG  484

Qy       172 SerThrMetPheGlySerValLeuSerTyrIleThrLeuArgLeuLeuGlyGluGlyPro  191
Db       485 TCCACAGTGTTTGGGACTGCCCTGAACTATGTGGCTCTCAGAATCCTGGGTATTGGGCCT  544

Qy       192 AsnAspGlyGlnGlyGluMetGluLysAlaArgAspTrpIleLeuGlyHisGlyGlyAla  211
Db       545 GATGAT------CCTGACCTTGTGCGCGCTCGGAACGTCCTTCACAAAAAAGGTGGTGCG  598

Qy       212 ThrTyrIleThrSerTrpGlyLysMetTrpLeuSerValLeuGlyValTyrGluTrpSer  231
Db       599 GTGGCCATCCCTTCCTGGGGGAAGTTCTGGCTGGCTGTCCTGAATGTTTATAGCTGGGAA  658

Qy       232 GlyAsnAsnProLeuProProGluIleTrpLeuLeuProTyrMetLeuProPheHisPro  251
Db       659 GGACTCAACACCCTATTCCCTGAGATGTGGCTGTTTCCTGAATGGGTACCAGCGCACCCC  718

Qy       252 GlyArgMetTrpCysHisCysArgMetValTyrLeuProMetSerTyrLeuTyrGlyLys  271
Db       719 TCCACTCTCTGGTGTCACTGCCGGCAGGTCTATCTGCCCATGAGTTACTGCTACGCCACT  778

Qy       272 ArgPheValGlyProIleSerProThrValLeuSerLeuArgLysGluLeuTyrThrVal  291
Db       779 CGGCTAAGTGCCTCAGAGGACCCACTGGTTCAGAGCCTCCGCCAGGAACTCTATGTGCAG  838

Qy       292 ProTyrHisAspIleAspTrpAspGlnAlaArgAsnLeuCysAlaLysGluAspLeuTyr  311
Db       839 GATTACGCCAGCATCGATTGGCCAGCGCAGAGGAACAACGTGTCCCCCGATGAGATGTAC  898

Qy       312 TyrProHisProLeuValGlnAspIleLeuTrpAlaSerLeuHisLysPheLeuGluPro  331
Db       899 ACGCCACACAGCTGGCTGCTGCATGTGGCGTATGGACTCCTCAAC--------------  943

Qy       332 IleLeuMetHisTrpProGlyLysArgLeuArgGluLysAlaIleIleSerAlaLeuGlu  351
Db       944 CTGTATGAACGTTTCCATAGCACCAGCCTGCGGAAGTGGGCCGTCCAGATGCTGTATGAA 1003

Qy       352 HisIleHisTyrGluAspGluAsnThrArgTyrIleCysIleGlyProValAsnLysVal  371
Db      1004 CACATTGCAGCTGATGACTGTTTCACTAAGTGCATCAGCATCGGCCCGATCTCAAAGACC 1063

Qy       372 LeuAsnMetLeuCysCysTrp---ValGluAspProAsnSerGluAlaPheLysLeuHis  390
Db      1064 ATCAACATGCTTGTTCGTTGGTCAGTGGATGGACCATCCTCCCCGGCCTTCCAGGAGCAC 1123

Qy       391 LeuProArgIleTyrAspTyrLeuTrpIleAlaGluAspGlyMetLysMetGlnGlyTyr  410
Db      1124 GTTTCTCGGATCAAAGATTACCTTTGGCTGGGCCTTGATGGCATGAAAATGCAGGGCACC 1183

Qy       411 AsnGlySerGlnLeuTrpAspThrAlaPheAlaValGlnAlaIleIleAlaSerAsnLeu  430
Db      1184 AATGGATCACAGATCTGGGACACTTCATTTGCTATCCAAGCACTGTTAGAGGCAGGTGCA 1243

Qy       431 IleGlu------GluPheGlyProThrIleArgLysAlaHisThrTyrIleLysAsnSer  448
Db      1244 CACCACAGACCTGAGTTTCTGCCCTGCCTGCAGAAGGCTCACGAGTTCCTGCGGCTTTCA 1303

Qy       449 GlnValLeuGluAspCysProGlyAspLeuAsnLysTrpTyrArgHisIleSerLysGly  468
Db      1304 CAGGTTCCAGAAAACTGTCCT---GACTATCAGAAGTACTATCGCCATATGCGCAAGGGT 1360

Qy       469 AlaTrpProPheSerThrGlyAspHisGlyTrpProIleSerAspCysThrAlaGluGly  488
Db      1361 GGTTTCTCCTTCAGCACACTGGACTGTGGCTGGATCGTTGCTGACTGCACAGCCGAGGGT 1420

Qy       489 LeuLysAlaValLeuLeuLeuSerLysIleAlaProGluIleValGlyGluProIleAsp  508
Db      1421 TTGAAGGCTGTGCTGCTCCTGCAGAATCAGTGTCCGTCCATCACT---GAGCATATCCCC 1477

Qy       509 ValLysArgLeuTyrAspSerValAsnValIleLeuSerLeuGlnAsnGluAspGlyGly  528
Db      1478 CGAGAGCGGCTGTGCGATGCTGTGGATGTGTTACTGAGCTTGAGGAATGCTGATGGCGGG 1537

Qy       529 PheAlaThrTyrGluLeuLysArgSerTyrAsnTrpLeuGluIleIleAsnProAlaGlu  548
Db      1538 TTCGCTACCTATGAGAAGAAGCGTGGGGGTTACTTGCTGGAGCTGCTGAACCCCTCAGAG 1597

Qy       549 ThrPheGlyAspIleValIleAspTyrProTyrValGluCysThrSerAlaAlaIleGln  568
Db      1598 GTCTTTGGAGACATCATGATTGACTACACGTATGTAGAGTGTACCTCCGCTGTGATGCAG 1657

Qy       569 AlaLeuAlaSerPheArgLysLeuTyrProGlyHisArgArgGluGluIleGlnHisCys  588
Db      1658 GCACTGAAGCATTTCCATGAACACTTCCCAGACTACAGGGCAGCAGAGGTCAGGGAGACC 1717

Qy       589 IleAspLysAlaThrThrPheIleGluLysIleGlnAlaSerAspGlySerTrpTyrGly  608
```

```
                ::::||||||   |||:::|||:::|||:::||||||||||:::|||   |||||||||
Db       1479 CGAGAGCGGCTGTGCGATGCTGTGGATGTGTTACTGAGCTTGAGGAATGCTGATGGCGGG 1538

Qy        529 PheAlaThrTyrGluLeuLysArgSerTyrAsnTrpLeuGluIleIleAsnProAlaGlu 548
              ||||||||||||||||   ||||||           ||||||:::::|||||:::|||
Db       1539 TTCGCTACCTATGAGAAGAAGCGTGGGGGTTACTTGCTGGAGCTGCTGAACCCCTCAGAG 1598

Qy        549 ThrPheGlyAspIleValIleAspTyrProTyrValGluCysThrSerAlaAlaIleGln 568
              ||||||||||||:::||||||||||   ||||||||||||||||||||   :::|||
Db       1599 GTCTTTGGAGACATCATGATTGACTACACGTATGTAGAGTGTACCTCCGCTGTGATGCAG 1658

Qy        569 AlaLeuAlaSerPheArgLysLeuTyrProGlyHisArgArgGluGluIleGlnHisCys 588
              ||||||      |||   :::   :::|||   :::|||      |||::::::
Db       1659 GCACTGAAGCATTTCCATGAACACTTCCCAGACTACAGGGCAGCAGAGGTCAGGGAGACC 1718

Qy        589 IleAspLysAlaThrThrPheIleGluLysIleGlnAlaSerAspGlySerTrpTyrGly 608
              :::::::::         |||       :::   |||   :::||||||||||||   |||
Db       1719 CTCAATCAAGGCCTGGACTTCTGCCGAAGGAAGCAGAGAGCCGATGGCTCCTGGGAGGGC 1778

Qy        609 SerTrpGlyValCysPheThrTyrGlyAlaTrpPheGlyValLysGlyLeuIleAlaAla 628
              ||||||||||||||||||||||||||   ||||||||:::::
Db       1779 TCCTGGGGTGTTTGTTTCACCTACGGCACCTGGTTTGGCTTGGAAGCGTTCGCTTGCATG 1838

Qy        629 GlyArgSerPheSerAsn---------CysSerSerIleArgLysAlaCysGluPheLeu 645
              |||   ::::::   :::         |||:::   :::   :::||||||   ||||||
Db       1839 GGGCATACCTACCAAGATGGGGCTGCTTGTGCAGAAGTAGCTCAGGCCTGCAACTTCCTC 1898

Qy        646 LeuSerLysGlnLeuProSerGlyGlyTrpGlyGluSerTyrLeuSerCysGlnAsnLys 665
              ||||||:::|||:::      ||||||||||||||   :::   ||||||:::   :::
Db       1899 CTGTCCCAGCAGATGGCGGATGGAGGCTGGGGAGAGGACTTTGAATCATGTGAACAGCGG 1958

Qy        666 ValTyrSerAsnLeuGluGlyAsnArgSerHisValValAsnThrGlyTrpAlaMetLeu 685
              |||       ::::::       ||||||   |||   :::|||   |||||::::::
Db       1959 CGGTAT------GTGCAGAGTGCTAGGTCCCAGGTCCATAGCACCTGCTGGGCCCTGATG 2012

Qy        686 AlaLeuIleAspAlaGlyGlnAlaLysArgAspSerGlnProLeuHisArgAlaAlaAla 705
              |||:::                 :::|||           |||
Db       2013 GGTTTGATGGCTGTCAGGCATCCCGACATCACTGCTCAG------GAGAGAGGGATCAGA 2066

Qy        706 TyrLeuIleAsnSerGlnLeuGluAspGlyAspPheProGlnGlnGluIleMetGlyVal 725
              |||:::      ||||||   :::|||||:::|||||::   |||   ||||||
Db       2067 TGTCTGCTAGGGAAGCAGTTACCCAATGGAGACTGGCCTCAGGAGAACATCTCTGGGGTC 2126

Qy        726 PheAsnLysAsnCysMetIleThrTyrAlaAlaTyrArgAsnIlePheProIleTrpAla 745
              |||||||||:::|||   |||:::|||   :::||||||||||||||||||||||||||
Db       2127 TTCAACAAGTCCTGTGCCATCAGCTACACAAGTTACAGAAACATCTTCCCCATCTGGGCC 2186

Qy        746 LeuGlyGluTyrGlnSer 751
              ||||||   :::   :::
Db       2187 CTGGGCCGCTTCTCCAAC 2204
```

RESULT 4
AK078023

```
LOCUS       AK078023               2836 bp    mRNA    linear   HTC 03-APR-2004
DEFINITION  Mus musculus 11 days embryo head cDNA, RIKEN full-length enriched
            library, clone:6230413B19 product:OXIDOSQUALENE CYCLASE homolog
            [Rattus norvegicus], full insert sequence.
ACCESSION   AK078023
VERSION     AK078023.1  GI:26346906
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 2836)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,
            URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.jp/
            URL:http://fantom.gsc.riken.jp/.
FEATURES             Location/Qualifiers
     source          1. .2836
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:6230413B19"
                     /db_xref="taxon:10090"
                     /clone="6230413B19"
                     /tissue_type="head"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="11 days embryo"
     CDS             38. .2239
                     /note="unnamed protein product; OXIDOSQUALENE CYCLASE
                     homolog [Rattus norvegicus] (SPTR|Q62811, evidence: FASTY,
                     92.1%ID, 100%length, match=2199)
```

```
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:A830080N10"
                     /db_xref="taxon:10090"
                     /clone="A830080N10"
                     /tissue_type="cortex"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="10 days neonate"
     CDS             39. .2240
                     /note="unnamed protein product; OXIDOSQUALENE CYCLASE
                     homolog [Rattus norvegicus] (SPTR|Q62811, evidence: FASTY,
                     92.1%ID, 100%length, match=2199)
                     putative"
                     /codon_start=1
                     /protein_id="BAC31739.1"
                     /db_xref="GI:26336108"
                     /translation="MTEGTCLRRRGGPYKTEPATDLTRWRLQNELGRQRWTYYQAEDD
                     PGREQTGLEAHSLGLDTRSYFTDLPKAQTAHEGALNGVTFYAKLQAEDGHWAGDYGGP
                     LPLLPGLLITCHISHISLPAGYREEMVRYLRSVQLPDGGWGLHIEDKSTVFGTALNYV
                     ALRILGIGPDDPDLVRARNVLHKKGGAVAIPSWGKFWLAVLNVYSWEGLNTLFPEMWL
                     FPEWVPAHPSTLWCHCRQVYLPMSYCYATRLSASEDPLVQSLRQELYVQDYASIDWPA
                     QRNNVSPDEMYTPHSWLLHVVYGLLNLYERFHSTSLRKWAVQMLYEHIAADDCFTKCI
                     SIGPISKTINMLARWSVDGPSSPAFQEHVSRIKDYLWLGLDGMKMQGTNGSQIWDTSF
                     AIQALLEAGAHHRPEFLPCLQKAHEFLRLSQVPENCPDYQKYYRHMRKGGFSFSTLDC
                     GWIVADCTAEGLKAVLLLQNQCPSITEHIPRERLCDAVDVLLSLRNADGGFATYEKKR
                     GGYLLELLNPSEVFGDIMIDYTYVECTSAVMQALKHFHEHFPDYRAAEVRETLNQGLD
                     FCRRKQRADGSWEGSWGVCFTYGTWFGLEAFACMGHTYQDGAACAEVAQACNFLLSQQ
                     MADGGWGEDFESCEQRRYVQSARSQVHSTCWALMGLMAVRHPDITAQERGIRCLLGKQ
                     LPNGDWPQENISGVFNKSCAISYTSYRNIFPIWALGRFSNLYPDNTLAGHI"
     polyA_signal    3075. .3080
                     /note="putative"
     polyA_site      3098
                     /note="putative"
ORIGIN

Alignment Scores:
Pred. No.:              6.73e-162       Length:         3098
Score:                  1670.50         Matches:        338
Percent Similarity:     61.39%          Conservative:   120
Best Local Similarity:  45.31%          Mismatches:     227
Query Match:            40.27%          Indels:         61
DB:                     3               Gaps:           15

US-09-419-679-6 (1-757) x AK044016 (1-3098)

Qy         17 LeuAsnAsnHisValGlyArgGlnValTrp----------GluPheAspProLys 31
Db        117 CTCCAGAATGAGTTGGGTCGGCAGAGATGGACTTATTATCAAGCGGAGGACGACCCT--- 173

Qy         32 LeuGlySerProGlnAspLeuLeuGluIleGluLysAlaArgGlnAsnPheHisAspAsn 51
Db        174 ---GGTCGAGAGCAGACCGGGCTAGAG---------------------------------- 197

Qy         52 ArgPheThrHisLysHisSerAlaAspLeuLeuMetArgMetGlnPheAlaArgGluAsn 71
Db        198 ------------GCCCACTCTTTGGGACTGGACACAAGAAGTTATTTCACGGAC------ 239

Qy         72 ProThrArgGluValLeuProLysValGlyValLysAspIleGluAspValThrGlnGlu 91
Db        240 --------------TTACCAAAAGCG----------------------------CAA 254

Qy         92 IleValThrLysThrLeuArgArgAlaValSerPheHisSerThrLeuGlnCysHisAsp 111
Db        255 ACAGCCCATGAGGGGGCCCTGAATGGAGTAACCTTTTATGCCAAGCTGCAGGCTGAGGAT 314

Qy        112 GlyHisTrpProGlyAspTyrGlyGlyProMetPheLeuMetProGlyLeuValIleThr 131
Db        315 GGACACTGGGCTGGTGATTATGGTGGTCCGCTCTTCCTCTTGCCAGGTCTCCTGATTACC 374

Qy        132 LeuSerIleThrGlyAlaLeuAsnThrValLeuThrGluGluHisArgLysGluIleCys 151
Db        375 TGTCACATATCA---------CACATCTCTCTGCCAGCTGGATACAGAGAGGAGATGGTG 425

Qy        152 ArgTyrLeuTyrAsnHisGlnAsnLysAspGlyGlyTrpGlyLeuHisIleGluGlyPro 171
Db        426 CGGTACCTGCGGTCAGTGCAGCTCCCTGATGGCGGCTGGGGCTTGCACATCGAGGACAAG 485

Qy        172 SerThrMetPheGlySerValLeuSerTyrIleThrLeuArgLeuLeuGlyGluGlyPro 191
Db        486 TCCACAGTGTTTGGGACTGCCCTGAACTATGTGGCTCTCAGAATCCTGGGTATTGGGCCT 545

Qy        192 AsnAspGlyGlnGlyGluMetGluLysAlaArgAspTrpIleLeuGlyHisGlyGlyAla 211
Db        546 GATGAT------CCTGACCTTGTGCGCGCTCGGAACGTCCTTCACAAAAAAGGTGGTGCG 599

Qy        212 ThrTyrIleThrSerTrpGlyLysMetTrpLeuSerValLeuGlyValTyrGluTrpSer 231
Db        600 GTGGCCATCCCTTCCTGGGGGAAGTTCTGGCTGGCTGTCCTGAATGTTTATAGCTGGGAA 659

Qy        232 GlyAsnAsnProLeuProProGluIleTrpLeuLeuProTyrMetLeuProPheHisPro 251
Db        660 GGACTCAACACCCTATTCCCTGAGATGTGGCTGTTTCCTGAATGGGTACCAGCGCACCCC 719

Qy        252 GlyArgMetTrpCysHisCysArgMetValTyrLeuProMetSerTyrLeuTyrGlyLys 271
Db        720 TCCACTCTCTGGTGTCACTGCCGGCAGGTCTATCTGCCCATGAGTTACTGCTACGCCACT 779

Qy        272 ArgPheValGlyProIleSerProThrValLeuSerLeuArgLysGluLeuTyrThrVal 291
Db        780 CGGCTAAGTGCCTCAGAGGACCCACTGGTTCAGAGCCTCCGCCAGGAACTCTATGTGCAG 839

Qy        292 ProTyrHisAspIleAspTrpAspGlnAlaArgAsnLeuCysAlaLysGluAspLeuTyr 311
Db        840 GATTACGCCAGCATCGATTGGCCAGCGCAGAGGAACAACGTGTCCCCCGATGAGATGTAC 899

Qy        312 TyrProHisProLeuValGlnAspIleLeuTrpAlaSerLeuHisLysPheLeuGluPro 331
Db        900 ACGCCACACAGCTGGCTGCTGCATGTGGTGTATGGACTCCTCAAC--------------- 944

Qy        332 IleLeuMetHisTrpProGlyLysArgLeuArgGluLysAlaIleIleSerAlaLeuGlu 351
Db        945 CTGTATGAACGTTTCCATAGCACCAGCCTGCGGAAGTGGGCCGTCCAGATGCTGTATGAA 1004

Qy        352 HisIleHisTyrGluAspGluAsnThrArgTyrIleCysIleGlyProValAsnLysVal 371
Db       1005 CACATTGCAGCTGATGACTGTTTCACTAAGTGCATCAGCATCGGCCCGATCTCAAAGACC 1064

Qy        372 LeuAsnMetLeuCysCysTrp---ValGluAspProAsnSerGluAlaPheLysLeuHis 390
Db       1065 ATCAACATGCTTGCTCGTTGGTCAGTGGATGGACCATCCTCCCCGGCCTTCCAGGAGCAC 1124

Qy        391 LeuProArgIleTyrAspTyrLeuTrpIleAlaGluAspGlyMetLysMetGlnGlyTyr 410
Db       1125 GTTTCTCGGATCAAAGATTACCTTTGGCTGGGCCTTGATGGCATGAAAATGCAGGGCACC 1184

Qy        411 AsnGlySerGlnLeuTrpAspThrAlaPheAlaValGlnAlaIleIleAlaSerAsnLeu 430
Db       1185 AATGGATCACAGATCTGGGACACTTCATTTGCTATCCAAGCACTGTTAGAGGCAGGTGCA 1244

Qy        431 IleGlu------GluPheGlyProThrIleArgLysAlaHisThrTyrIleLysAsnSer 448
Db       1245 CACCACAGACCTGAGTTTCTGCCCTGCCTGCAGAAGGCTCACGAGTTCCTGCGGCTTTCA 1304

Qy        449 GlnValLeuGluAspCysProGlyAspLeuAsnLysTrpTyrArgHisIleSerLysGly 468
Db       1305 CAGGTTCCAGAAAACTGTCCT---GACTATCAGAAGTACTATCGCCATATGCGCAAGGGT 1361

Qy        469 AlaTrpProPheSerThrGlyAspHisGlyTrpProIleSerAspCysThrAlaGluGly 488
Db       1362 GGTTTCTCCTTCAGCACACTGGACTGTGGCTGGATCGTTGCTGACTGCACAGCCGAGGGT 1421

Qy        489 LeuLysAlaValLeuLeuLeuSerLysIleAlaProGluIleValGlyGluProIleAsp 508
Db       1422 TTGAAGGCTGTGCTGCTCCTGCAGAATCAGTGTCCGTCCATCACT---GAGCATATCCCC 1478

Qy        509 ValLysArgLeuTyrAspSerValAsnValIleLeuSerLeuGlnAsnGluAspGlyGly 528
```

```
Db       1258 GTTCAAGCTATAGTGGCTACTGGCCTCATTGAAGAATTTGGTCCTACTCTTAAACTAGCA 1317

Qy        442 HisThrTyrIleLysAsnSerGlnValLeuGluAspCysProGlyAspLeuAsnLysTrp 461

Db       1318 CATGGCTACATAAAGAAAACGCAGGTTATCGATGACTGCCCTGGAGATCTTAGTCAGTGG 1377

Qy        462 TyrArgHisIleSerLysGlyAlaTrpProPheSerThrGlyAspHisGlyTrpProIle 481

Db       1378 TACCGCCACATATCTAAAGGTGCATGGCCCTTTTCTACTGCTGATCATGGTTGGCCTATA 1437

Qy        482 SerAspCysThrAlaGluGlyLeuLysAlaValLeuLeuLeuSerLysIleAlaProGlu 501

Db       1438 TCAGATTGCACTGCAGAAGGACTTAAGGCGGCATTATTGCTATCGAAGATTTCTCCAGAT 1497

Qy        502 IleValGlyGluProIleAspValLysArgLeuTyrAspSerValAsnValIleLeuSer 521

Db       1498 ATTGTTGGCGAAGCAGTGGAAGTTAATAGACTGTATGATTCTGTCAATTGTTTGATGTCA 1557

Qy        522 LeuGlnAsnGluAspGlyGlyPheAlaThrTyrGluLeuLysArgSerTyrAsnTrpLeu 541

Db       1558 TACATGAATGATAATGGTGGATTTGCAACATATGAACTCACAAGGTCTTATGCCTGGCTG 1617

Qy        542 GluIleIleAsnProAlaGluThrPheGlyAspIleValIleAspTyrProTyrValGlu 561

Db       1618 GAGCTTATCAATCCTGCTGAGACCTTTGGGGACATTGTGATTGATTATCCTTATGTGGAA 1677

Qy        562 CysThrSerAlaAlaIleGlnAlaLeuAlaSerPheArgLysLeuTyrProGlyHisArg 581

Db       1678 TGCACTTCAGCAGCAATTCAGGCCCTGACAGCATTTAAAAAGCTCTACCCTGGACACCGC 1737

Qy        582 ArgGluGluIleGlnHisCysIleAspLysAlaThrThrPheIleGluLysIleGlnAla 601

Db       1738 AAGAGTGAAATAGACAACTGTATAAGCAAAGCTGCTAGCTTTATTGAGGGTATTCAAAAA 1797

Qy        602 SerAspGlySerTrpTyrGlySerTrpGlyValCysPheThrTyrGlyAlaTrpPheGly 621

Db       1798 AGCGATGGTTCATGGTATGGTTCTTGGGCTGTTTGTTTTACCTATGGCACATGGTTTGGT 1857

Qy        622 ValLysGlyLeuIleAlaAlaGlyArgSerPheSerAsnCysSerSerIleArgLysAla 641

Db       1858 GTAAAGGGATTAGTTGCTGCTGGTAGGACATTCAAAAACAGTCCTGCAATCAGAAAGGCA 1917

Qy        642 CysGluPheLeuLeuSerLysGlnLeuProSerGlyGlyTrpGlyGluSerTyrLeuSer 661

Db       1918 TGTGACTTTTTGTTGTCAAAAGAGCTTCCTTCTGGAGGCTGGGGAGAAAGCTATTTGTCA 1977

Qy        662 CysGlnAsnLysValTyrSerAsnLeuGluGlyAsnArgSerHisValValAsnThrGly 681

Db       1978 TCCCAAGATCAGGTTTATACCAATCTCGAAGGGAAGCGACCTCATGCGGTGAACACTGGT 2037

Qy        682 TrpAlaMetLeuAlaLeuIleAspAlaGlyGlnAlaLysArgAspSerGlnProLeuHis 701

Db       2038 TGGGCCATGCTAGCCCTAATCGATGCAGGGCAGGCTGAGAGAGATCCAATTCCTTTGCAT 2097

Qy        702 ArgAlaAlaAlaTyrLeuIleAsnSerGlnLeuGluAspGlyAspPheProGlnGlnGlu 721

Db       2098 CGAGCAGCGAAGGTTTTGATCAACTTACAATCGGAAGATGGTGAATTTCCCCAGCAAAGC 2157

Qy        722 Ile 722

Db       2158 ATT 2160

RESULT 3
AK044016
LOCUS       AK044016               3098 bp    mRNA    linear   HTC 03-APR-2004
DEFINITION  Mus musculus 10 days neonate cortex cDNA, RIKEN full-length
            enriched library, clone:A830080N10 product:OXIDOSQUALENE CYCLASE
            homolog [Rattus norvegicus], full insert sequence.
ACCESSION   AK044016
VERSION     AK044016.1  GI:26336107
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 3098)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,
            URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.jp/
            URL:http://fantom.gsc.riken.jp/.
FEATURES             Location/Qualifiers
     source          1. .3098
                     /organism="Mus musculus"
```

```
Qy        634 AsnCysSerSerIleArgLysAlaCysGluPheLeuLeuSerLysGlnLeuProSerGly 653
              |||      :::|||||||||||||||||:::|||||||||||||||:::||||||||||||
Db       2138 AACAGTCCTGCAATTAGAAAGGCATGCGACTTTCTGTTGTCAAAAGAACTTCCTTCCGGT 2197

Qy        654 GlyTrpGlyGluSerTyrLeuSerCysGlnAsnLysValTyrSerAsnLeuGluGlyAsn 673
              |||||||||||||||||||||||   |||::::::|||||:::||||||:::||||||
Db       2198 GGTTGGGGAGAAAGCTATTTGTCATCTCAAGACCAGGTTTACACCAATCTCAAAGGCAAC 2257

Qy        674 ArgSerHisValValAsnThrGlyTrpAlaMetLeuAlaLeuIleAspAlaGlyGlnAla 693
              |||  |||  ||||||||  ||||||||||||||||||||||||||||||||||||||||
Db       2258 CGGCCTCATGCGGTGAACACTAGTTGGGCCATGCTGGCGCTGATTGATGCGGGCCAGGCC 2317

Qy        694 LysArgAspSerGlnProLeuHisArgAlaAlaAlaTyrLeuIleAsnSerGlnLeuGlu 713
              :::||||||      ||||||||||||||||||      |||||||||  |||  |||
Db       2318 GAGAGAGATCCAACGCCTCTACACCGAGCAGCAAAGGTTTTGATCAACTTACAATCAGAG 2377

Qy        714 AspGlyAspPheProGlnGlnGluIleMetGlyValPheAsnLysAsnCysMetIleThr 733
              ||||||:::|||||||||||||||||||:::|||||||||||||||||||||||||||:::
Db       2378 GACGGAGAATTTCCTCAGCAAGAGATCATAGGAGTGTTCAACAAGAACTGCATGATAAGC 2437

Qy        734 TyrAlaAlaTyrArgAsnIlePheProIleTrpAlaLeuGlyGluTyrGlnSerGlnVal 753
              |||:::  ||||||||||||||||||||||||||||||||||||||:::   :::|||
Db       2438 TACTCCCAGTACAGGAACATCTTCCCGATTTGGGCTCTGGGTGAGTACCGGTGTCGAGTC 2497
```

RESULT 2
CL965443
LOCUS CL965443 2166 bp DNA linear GSS 21-SEP-2004
DEFINITION OsIFCC012474 Oryza sativa Express Library Oryza sativa (indica cultivar-group) genomic, genomic survey sequence.
ACCESSION CL965443
VERSION CL965443.1 GI:52385571
KEYWORDS GSS.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 2166)
AUTHORS Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M., Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L., Wong,G.K.S., Deng,X.W. and Wang,J.
TITLE An analysis of transcriptional regulation of the rice genome and its comparison to Arabidopsis
JOURNAL Unpublished (2004)
COMMENT Contact: Chen Chen
Department of Bioinformatic
Beijing Institute of Genomics
Chinese Academy of Sciences, Beijing 101300, China
Tel: 86-10-80481559
Fax: 86-10-80488676
Email: chenchen@genomics.org.cn
Rice genomic sequence.
Class: exon-trapped.
FEATURES Location/Qualifiers
source 1. .2166
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="genomic DNA"
/db_xref="taxon:39946"
/clone_lib="Oryza sativa Express Library"
/note="Oryza sativa exon trapped genomic sequences "
ORIGIN

Alignment Scores:

Pred. No.:	2.6e-289	Length:	2166
Score:	2891.00	Matches:	499
Percent Similarity:	87.90%	Conservative:	82
Best Local Similarity:	75.49%	Mismatches:	80
Query Match:	69.70%	Indels:	0
DB:	9	Gaps:	0

US-09-419-679-6 (1-757) x CL965443 (1-2166)

```
Qy         62 LeuMetArgMetGlnPheAlaArgGluAsnProThrArgGluValLeuProLysValGly 81
              :::  |||  ||||||||||:::  ||||||  :::      :::|||  :::
Db        178 ATGCAGAGAACTCAGTTTGCTAAAGCAAATCCTCTTAAGCTGGACATTCCAGCTATCAAG 237

Qy         82 ValLysAspIleGluAspValThrGlnGluIleValThrLysThrLeuArgArgAlaVal 101
              :::::::::  |||  ||||||  |||  |||      :::|||:::||||||:::
Db        238 CTAGAGGAGCATGAAGCTGTTACAGGCGAAGCTGTATTATCCTCTCTAAAAAGGGCAATT 297

Qy        102 SerPheHisSerThrLeuGlnCysHisAspGlyHisTrpProGlyAspTyrGlyGlyPro 121
              :::  :::||||||  |||  ||||||||||||||||||||||||||||||||||||||
Db        298 GCTCGTTATTCTACTTTTCAAGCACATGATGGACACTGGCCTGGGGATTATGGTGGTCCC 357

Qy        122 MetPheLeuMetProGlyLeuValIleThrLeuSerIleThrGlyAlaLeuAsnThrVal 141
              ||||||||||||||||||:::||||||||  ::::::||||||||||||||
Db        358 ATGTTCCTTATGCCAGGCTTGATTATAACATTATATGTGAGCGGAGCGCTGAACACTGCT 417

Qy        142 LeuThrGluGluHisArgLysGluIleCysArgTyrLeuTyrAsnHisGlnAsnLysAsp 161
              |||:::  ||||||:::||||||||  ||||||||||||||||||||||:::|||
Db        418 TTGTCATCTGAACATCAGAAGGAGATACGGCGGTATCTCTACAATCATCAGAATGAAGAT 477

Qy        162 GlyGlyTrpGlyLeuHisIleGluGlyProSerThrMetPheGlySerValLeuSerTyr 181
              ||||||||||||||||||||||||  ||||||||||||||||  |||:::|||
Db        478 GGAGGATGGGGTTTGCACATTGAGGGCCACAGCACCATGTTTGGCTCAGCGTTGACATAT 537

Qy        182 IleThrLeuArgLeuLeuGlyGluGlyProAsnAspGlyGlnGlyGluMetGluLysAla 201
              ::::::|||||||||||||||||||:::  |||  |||  |||||||||
Db        538 GTTTCTTTGAGATTGCTTGGGGAGGGGCCAGATAGCGGCGATGGAGCTATGGAGAAAGGA 597

Qy        202 ArgAspTrpIleLeuGlyHisGlyGlyAlaThrTyrIleThrSerTrpGlyLysMetTrp 221
              |||  ||||||||  ||||||||||||||||||||||||||||||||  |||
Db        598 CGAAAATGGATTTTAGACCATGGGGGAGCAACTTATATTACATCATGGGGCAAGTTTTGG 657

Qy        222 LeuSerValLeuGlyValTyrGluTrpSerGlyAsnAsnProLeuProProGluIleTrp 241
              ||||||||||||||||||::::::|||||||||||||||||:::|||||||||||||||
Db        658 CTTTCGGTGCTTGGTGTATTTGACTGGTCTGGCAACAACCCAGTGCCACCAGAAATATGG 717

Qy        242 LeuLeuProTyrMetLeuProPheHisProGlyArgMetTrpCysHisCysArgMetVal 261
              ||||||||||||  ||||||  ||||||||||||||||||||||||||||||||||||||
Db        718 TTGTTGCCATATTTCCTGCCGATTCATCCAGGGCGAATGTGGTGTCATTGCCGGATGGTT 777

Qy        262 TyrLeuProMetSerTyrLeuTyrGlyLysArgPheValGlyProIleSerProThrVal 281
              ||||||||||||  |||:::||||||||||||||||||||||||:::::::|||  :::
Db        778 TATTTGCCTATGTGTTACATTTATGGAAAGAGGTTTGTGGGCCCAGTTACACCAATTATA 837

Qy        282 LeuSerLeuArgLysGluLeuTyrThrValProTyrHisAspIleAspTrpAspGlnAla 301
              |||  ||||||||||||||||||  ||||||||:::::::::|||||||:::|||
Db        838 TTGGAATTAAGAAAGGAACTCTACGAAGTACCCTACAATGAAGTTGATTGGGACAAGGCT 897

Qy        302 ArgAsnLeuCysAlaLysGluAspLeuTyrTyrProHisProLeuValGlnAspIleLeu 321
              ||||||||||||||||||||||||||||||||||||||||||  :::||||||:::|||
Db        898 CGCAATCTATGTGCTAAGGAAGATCTGTACTATCCACATCCATTCATGCAGGATGTATTA 957

Qy        322 TrpAlaSerLeuHisLysPheLeuGluProIleLeuMetHisTrpProGlyLysArgLeu 341
              ||||||:::|||||||||||||:::||||||  ::::::  ||||||||  :::|||
Db        958 TGGGCCACTCTCCACAAATTTGTTGAACCAGCTATGTTGCGTTGGCCTGGGAACAAATTG 1017

Qy        342 ArgGluLysAlaIleIleSerAlaLeuGluHisIleHisTyrGluAspGluAsnThrArg 361
              ||||||||||||:::  :::  ::::::||||||||||||||||||||||||||||||
Db       1018 AGGGAGAAAGCTTTGGACACTGTCATGCAGCATATTCATTATGAAGATGAGAACACCCGA 1077

Qy        362 TyrIleCysIleGlyProValAsnLysValLeuAsnMetLeuCysCysTrpValGluAsp 381
              ||||||||||||||||||||||||||||||||||||||||||||  ||||||:::||||||
Db       1078 TATATTTGCATTGGTCCAGTAAACAAGGTATTAAATATGCTTGCTTGCTGGATTGAAGAT 1137

Qy        382 ProAsnSerGluAlaPheLysLeuHisLeuProArgIleTyrAspTyrLeuTrpIleAla 401
              ||||||||||||||||||||||||:::|||||:::::||||||||||||||||||||||
Db       1138 CCAAACTCAGAGGCATTCAAACTCCACATTCCAAGAGTCCACGATTACCTATGGATTGCA 1197

Qy        402 GluAspGlyMetLysMetGlnGlyTyrAsnGlySerGlnLeuTrpAspThrAlaPheAla 421
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1198 GAAGATGGCATGAAAATGCAGGGTTATAATGGAAGCCAGCTGTGGGACACAGCTTTCACA 1257

Qy        422 ValGlnAlaIleIleAlaSerAsnLeuIleGluGluPheGlyProThrIleArgLysAla 441
```

/mol_type="mRNA"
/db_xref="MaizeDB:630740"
/db_xref="taxon:4577"
/clone_lib="Maize Mapping Project/DuPont Cornsensus Library"
/note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed DuPont contigs; this resource was assembled by DuPont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"

ORIGIN

Alignment Scores:

Pred. No.:	4.75e-304	Length:	2920
Score:	3034.50	Matches:	543
Percent Similarity:	83.11%	Conservative:	72
Best Local Similarity:	73.38%	Mismatches:	120
Query Match:	73.16%	Indels:	5
DB:	3	Gaps:	1

US-09-419-679-6 (1-757) x AY109591 (1-2920)

```
Qy         14 LeuArgThrLeuAsnAsnHisValGlyArgGlnValTrpGluPheAspProLysLeuGly 33
              |||||||||   |||   |||||||||||||||||||||||||||
Db        293 CTCCGCACGAAGAACGCCCACGTCGGCCGCCAGGTCTGGGAGTTC---------------- 337

Qy         34 SerProGlnAspLeuLeuGluIleGluLysAlaArgGlnAsnPheHisAspAsnArgPhe 53

Db        338 GNNNNNNNNNNNNNNNNNNNNNNNNNNNCGNNNNNNNNNNNNNNNNNNNNNNNNNNGC 397

Qy         54 ThrHisLysHisSerAlaAspLeuLeuMetArgMetGlnPheAlaArgGluAsnProThr 73
              |||   |||||||||||||||||||||||||||:::||||||||||:::||||||
Db        398 ACGAGCAAGCACAGCGCCGACCTCCTCATGCGCATCCAGTTCGCCAAAGAAAACTCGATT 457

Qy         74 ArgGluValLeuProLysValGlyValLysAspIleGluAspValThrGlnGluIleVal 93
                      ||||||   :::   :::   :::   ||||||||||||:::|||   |||
Db        458 GAGCTTCACCTTCCAGGCATCAAGCTCGGTGAGCATGAAGATGTGACCGAGGAAGCTGTG 517

Qy         94 ThrLysThrLeuArgArgAlaValSerPheHisSerThrLeuGlnCysHisAspGlyHis 113
                    |||||||:::||||||:::|||       ||||||||||   ||||||||||||
Db        518 TTGACTACATTGAAAAGGGCAATCAGCCGTTTCTCTACTCTCCAGGCACATGATGGACAC 577

Qy        114 TrpProGlyAspTyrGlyGlyProMetPheLeuMetProGlyLeuValIleThrLeuSer 133
              |||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db        578 TGGCCTGGGGATTATGGTGGTCCTATGTTCCTTATGCCAGGCTTGATCATAACATTGTAT 637

Qy        134 IleThrGlyAlaLeuAsnThrValLeuThrGluGluHisArgLysGluIleCysArgTyr 153
              :::|||||||||||||||||||||||:::   ||||||:::|||||||||   ||||||
Db        638 GTGACTGGAGCACTAAACACTGTCTTGTCATTGGAACACCAGAAGGAGATCCGCCGGTAT 697

Qy        154 LeuTyrAsnHisGlnAsnLysAspGlyGlyTrpGlyLeuHisIleGluGlyProSerThr 173
              ||||||||||||||||||:::||||||||||||||||||||||||||||||||||||||
Db        698 CTTTATAATCACCAGAATGAAGATGGCGGCTGGGGCTTGCACATTGAGGGTCCAAGCACC 757

Qy        174 MetPheGlySerValLeuSerTyrIleThrLeuArgLeuLeuGlyGluGlyProAsnAsp 193
              ||||||||||||   |||:::|||:::   |||||||||||||||||||||||||:::
Db        758 ATGTTCGGCTCAGCACTGACCTATGTTATTTTGAGATTGCTTGGAGAGGGACCAGATAGT 817

Qy        194 GlyGlnGlyGluMetGluLysAlaArgAspTrpIleLeuGlyHisGlyGlyAlaThrTyr 213
              |||   |||   |||||||||   |||:::|||||||||   ||||||||||||||||||
Db        818 GGAGATGGAGCCATGGAGAAAGGTCGAAACTGGATATTAGACCATGGTGGAGCAACATAT 877

Qy        214 IleThrSerTrpGlyLysMetTrpLeuSerValLeuGlyValTyrGluTrpSerGlyAsn 233
              |||||||||      |||   |||||||||||||||||||:::|||||||||||||||
Db        878 ATAACATCGTNNNNNAAGTTTTGGCTTTCGGTACTAGGTGTATTTGAATGGTCTGGTAAC 937

Qy        234 AsnProLeuProProGluIleTrpLeuLeuProTyrMetLeuProPheHisProGlyArg 253
              ||||||:::|||||||||:::|||||||||||||||:::|||||||||||||||||||||
Db        938 AACCCGGTGCCACCAGAAGTATGGCTACTGCCATATCTCCTCCCATTTCACCCAGGGAGG 997

Qy        254 MetTrpCysHisCysArgMetValTyrLeuProMetSerTyrLeuTyrGlyLysArgPhe 273
              ||||||||||||||||||||||||||||||||||   |||:::|||||||||||||||||
```

```
Db        998 ATGTGGTGTCATTGTCGAATGGTGTATTTGCCAATGTGCTACATTTATGGGAAGAGGTTT 1057

Qy        274 ValGlyProIleSerProThrValLeuSerLeuArgLysGluLeuTyrThrValProTyr 293
              ||||||   |||:::|||   :::|||   ||||||||||||||:::      ||||||
Db       1058 GTTGGCCGAATCACACCACTTCTGTTGGAATTAAGGAAGGAACTTTTCAAAGACCCTTAC 1117

Qy        294 HisAspIleAspTrpAspGlnAlaArgAsnLeuCysAlaLysGluAspLeuTyrTyrPro 313
                    ||||||||||||:::||||||||||||||||||||||||||||||||||||||
Db       1118 AGCAAGATTGATTGGGACAAGGCCCGCAACCTATGTGCCAAGGAAGATCTGTACTACCCA 1177

Qy        314 HisProLeuValGlnAspIleLeuTrpAlaSerLeuHisLysPheLeuGluProIleLeu 333
              ||||||   ||||||||:::||||||||:::||||||||||||:::||||||::::::
Db       1178 CACCCATTCGTTCAAGATGTGTTGTGGGCCACTCTCCATAAATTCGTTGAACCAGTTATG 1237

Qy        334 MetHisTrpProGlyLysArgLeuArgGluLysAlaIleIleSerAlaLeuGluHisIle 353
              ||||||||||||||   :::||||||||||||||:::   :::   ::::::|||:::
Db       1238 ATGCATTGGCCTGGCAGCAAATTGAGGGAGAAAGCTCTGGAAACAGTCATGCAACATGTT 1297

Qy        354 HisTyrGluAspGluAsnThrArgTyrIleCysIleGlyProValAsnLysValLeuAsn 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1298 CATTATGAAGATGAGAACACTCGTTATATTTGCATTGGTCCTGTAAACAAGGTATTGAAT 1357

Qy        374 MetLeuCysCysTrpValGluAspProAsnSerGluAlaPheLysLeuHisLeuProArg 393
                           |||||||||||||||||||||||||||||||:::||||||
Db       1358 ANNNNNNNNNNNNNNNNNNAAGATCCAAACTCGGAGGCCTTCAAACTTCATATCCCACGA 1417

Qy        394 IleTyrAspTyrLeuTrpIleAlaGluAspGlyMetLysMetGlnGlyTyrAsnGlySer 413
              :::||||||||||||:::||||||||||||||||||||||||||||||||||||||||||
Db       1418 GTCTATGATTACTTGTGGCTTGCTGAAGATGGCATGAAGATGCAGGGTTATAATGGCAGC 1477

Qy        414 GlnLeuTrpAspThrAlaPheAlaValGlnAlaIleIleAlaSerAsnLeuIleGluGlu 433
              ||||||||||||||||||||   ||||||||||||:::|||:::||||||||||||||||
Db       1478 CAACTTTGGGATACAGCTTTCACAGTTCAAGCCATTGTGGCTACCAACCTTATTGAAGAG 1537

Qy        434 PheGlyProThrIleArgLysAlaHisThrTyrIleLysAsnSerGlnValLeuGluAsp 453
              ||||||||||||::::::   ||||||   ||||||||||||||||||||||||:::|||
Db       1538 TTTGGTCCTACCCTTAAACTAGCACACAACTATATCAAGAATTCACAGGTTCTTGATGAC 1597

Qy        454 CysProGlyAspLeuAsnLysTrpTyrArgHisIleSerLysGlyAlaTrpProPheSer 473
              ||||||||||||||||||   ||||||||||   ||||||||||||||||||||||||||
Db       1598 TGCCCTGGTGATCTGAATGACTGGTACCGCCACACATCTAAAGGTGCATGGCCATTCTCA 1657

Qy        474 ThrGlyAspHisGlyTrpProIleSerAspCysThrAlaGluGlyLeuLysAlaValLeu 493
              |||   ||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db       1658 ACTGCTGATCATGGTTGGCCTATATCTGATTGCACTGCTGAAGGACTAAAGGCTTCATTA 1717

Qy        494 LeuLeuSerLysIleAlaProGluIleValGlyGluProIleAspValLysArgLeuTyr 513
              |||||||||:::|||                                  |||   |||
Db       1718 TTGTTATCAAGGATCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGCTAATAGATTTTAT 1777

Qy        514 AspSerValAsnValIleLeuSerLeuGlnAsnGluAspGlyGlyPheAlaThrTyrGlu 533
              |||:::|||:::   ::::::|||       |||::::::||||||||||||||||||||
Db       1778 GATGCTGTCAGTTGTCTGATGTCTTATATGAATGATAATGGCGGTTTCGCGACATATGAA 1837

Qy        534 LeuLysArgSerTyrAsnTrpLeuGluIleIleAsnProAlaGluThrPheGlyAspIle 553
              |||   |||||||||   ||||||||||:::||||||||||||||||||||||||||||
Db       1838 CTCACAAGATCTTATCCCTGGTTGGAGCTGATCAATCCCGCTGAGACCTTTGGGGATATT 1897

Qy        554 ValIleAspTyrProTyrValGluCysThrSerAlaAlaIleGlnAlaLeuAlaSerPhe 573
              ||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db       1898 GTGATTGATTACCCGTATGTTGAATGTACATCAGCAGCAATTCAGGCCCTGACATCATTC 1957

Qy        574 ArgLysLeuTyrProGlyHisArgArgGluGluIleGlnHisCysIleAspLysAlaThr 593
                    ||||||||||||||||||||:::|||:::   :::||||||   ||||||:::
Db       1958 NNNNNNCTATACCCTGGGCACCGCAGGAAAGAGGTGGATAACTGTATCAGCAAAGCTTCC 2017

Qy        594 ThrPheIleGluLysIleGlnAlaSerAspGlySerTrpTyrGlySerTrpGlyValCys 613
                  |||||||||   ||||||   ||||||||||||||||||||||||||   ||||||
Db       2018 AATTTCATCGAGAGTATTCAGAAAAGCGATGGTTCATGGTATGGCTCTTGGGCCGTCTGT 2077

Qy        614 PheThrTyrGlyAlaTrpPheGlyValLysGlyLeuIleAlaAlaGlyArgSerPheSer 633
              ||||||||||||   ||||||||||||||||||||||||||||||||||:::|||
Db       2078 TTCACATACGGCACTTGGTTTGGTGTGAAGGGACTAATTGCTGCTGGTAGAACATTTGAG 2137
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 5, 2005, 12:09:54 ; Search time 6115 Seconds
(without alignments)
4712.126 Million cell updates/sec

Title: US-09-419-679-6
Perfect score: 4148
Sequence: 1 MWKLKFAEGGNPWLRTLNNH.....RNIFPIWALGEYQSQVLQSR 757

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/6630617/runat_02112005_120301_15089/app_query.fasta_1.903
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=6630617_@CGN_1_1_3437_@runat_02112005_120301_15089 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_htc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3034.5	73.2	2920	3	AY109591	AY109591 Zea mays
2	2891	69.7	2166	9	CL965443	CL965443 OsIFCC012
3	1670.5	40.3	3098	3	AK044016	AK044016 Mus muscu
4	1667.5	40.2	2836	3	AK078023	AK078023 Mus muscu
5	1653.5	39.9	4659	3	AK031271	AK031271 Mus muscu
6	1652	39.8	1887	9	CL964361	CL964361 OsIFCC010
7	1407	33.9	943	7	CK265941	CK265941 EST712019
8	1374	33.1	859	4	BG588039	BG588039 EST489814
9	1309	31.6	958	7	CK251308	CK251308 EST734945
10	1299	31.3	850	6	CD440985	CD440985 EL01N0563
11	1257	30.3	840	6	CB976421	CB976421 CAB40002_
12	1253	30.2	831	6	CB642220	CB642220 OSJNEb02E
13	1242	29.9	803	6	CB673111	CB673111 OSJNEe07H
14	1222	29.5	781	6	CB651302	CB651302 OSJNEb16D
15	1210	29.2	689	7	CK768972	CK768972 Gm-r1030-
16	1206	29.1	768	6	CB622107	CB622107 OSIIEa08H
17	1171	28.2	809	4	BG414397	BG414397 HVSMEk000
18	1169	28.2	707	5	BQ866177	BQ866177 QGC7E04.y
19	1149	27.7	638	5	BU926899	BU926899 sas93a11.
20	1145	27.6	848	6	CB620259	CB620259 OSIIEa05F
21	1128	27.2	708	7	CV518867	CV518867 0089P0020
22	1124	27.1	666	2	BE330835	BE330835 so86a10.y
c 23	1115	26.9	729	2	BE821083	BE821083 GM700014A
24	1112.5	26.8	770	4	BG595048	BG595048 EST493726
c 25	1099	26.5	824	7	CO984997	CO984997 GM89023A1
26	1098	26.5	673	7	CV095432	CV095432 FAMU_USDA
27	1098	26.5	760	4	BI920796	BI920796 EST540719
28	1090	26.3	3971	3	AK054393	AK054393 Mus muscu
29	1088	26.2	703	6	CA160874	CA160874 SCMCRZ306
30	1088	26.2	756	4	BG648956	BG648956 EST510575
31	1081	26.1	626	2	AW624112	AW624112 EST322057
32	1069	25.8	681	6	CD006346	CD006346 VVB058B07
33	1055	25.4	860	6	CB642442	CB642442 OSJNEb02K
34	1054	25.4	664	5	BU040703	BU040703 PP_LEa000
35	1053	25.4	672	6	CD714862	CD714862 VVB205F07
c 36	1052	25.4	901	7	CV227637	CV227637 WS0167.B2
37	1051.5	25.3	701	7	CF448640	CF448640 EST684985
38	1044	25.2	651	6	CD890082	CD890082 G118.113N
39	1044	25.2	666	6	CD718769	CD718769 VVB153H11
40	1032	24.9	662	6	CD714635	CD714635 VVB203B10
41	1031	24.9	674	4	BM437263	BM437263 VVA016H02
42	1029	24.8	719	6	CA125303	CA125303 SCRFLR105
43	1027	24.8	627	7	CF303833	CF303833 ABF1--03-
44	1025	24.7	642	6	CD010811	CD010811 VVB115H08
45	1024	24.7	593	2	AW931505	AW931505 EST357348

ALIGNMENTS

RESULT 1
AY109591
LOCUS AY109591 2920 bp mRNA linear HTC 17-OCT-2002
DEFINITION Zea mays CL201_1 mRNA sequence.
ACCESSION AY109591
VERSION AY109591.1 GI:21213371
KEYWORDS HTC.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 2920)
AUTHORS Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S., Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes
JOURNAL Unpublished (2002)
REFERENCE 2 (bases 1 to 2920)
AUTHORS Coe,E.H.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA
COMMENT If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmDB: www.zmdb.iastate.edu.
FEATURES Location/Qualifiers
source 1. .2920
/organism="Zea mays"

THIS PAGE BLANK (USPTO)

```
Qy        678 ValAsnThrGlyTrpAlaMetLeuAlaLeuIleAspAlaGlyGlnAlaLysArgAspSer 697
              |||:::|||   ||||||::::::   ||||||   ||||||||||||   ||||||
Db       2071 GTACATACAGCATGGGCTCTTATGGGACTAATTCATGCTGGACAGGCGGATAGAGACCCC 2130

Qy        698 GlnProLeuHisArgAlaAlaAlaTyrLeuIleAsnSerGlnLeuGluAspGlyAspPhe 717
                ||||||||||||||||||      ||||||||||||||||||||:::||||||:::
Db       2131 ATGCCTCTTCACCGTGCTGCAAAGTTGCTCATTAATTCTCAGTTGGAAGAGGGTGATTGG 2190

Qy        718 ProGlnGlnGluIleMetGlyValPheAsnLysAsnCysMetIleThrTyrAlaAlaTyr 737
              ||||||||||||||   |||||||||   ||||||||||:::   |||       |||
Db       2191 CCCCAACAGGAAATCACGGGAGTATTCATGAAAAATTGCATGTTGCATTATCCAATGTAC 2250

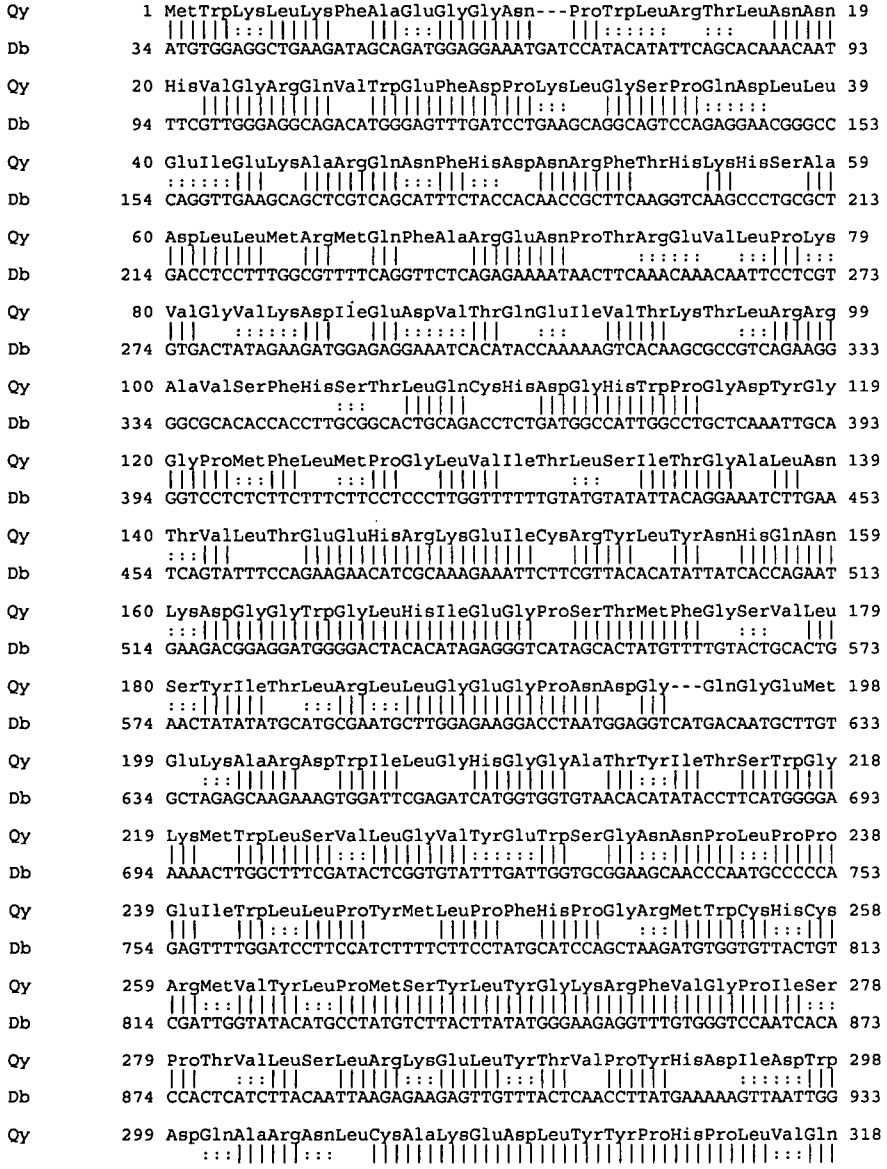
Qy        738 ArgAsnIleProIleTrpAlaLeuGlyGluTyrGlnSerGlnVal 753
              |||:::|||:::|||:::|||||||||   ||||||:::   :::|||
Db       2251 AGAGATATTTATCCAATGTGGGCTCTAGCTGAATATCGAAGGCGGGTT 2298
```

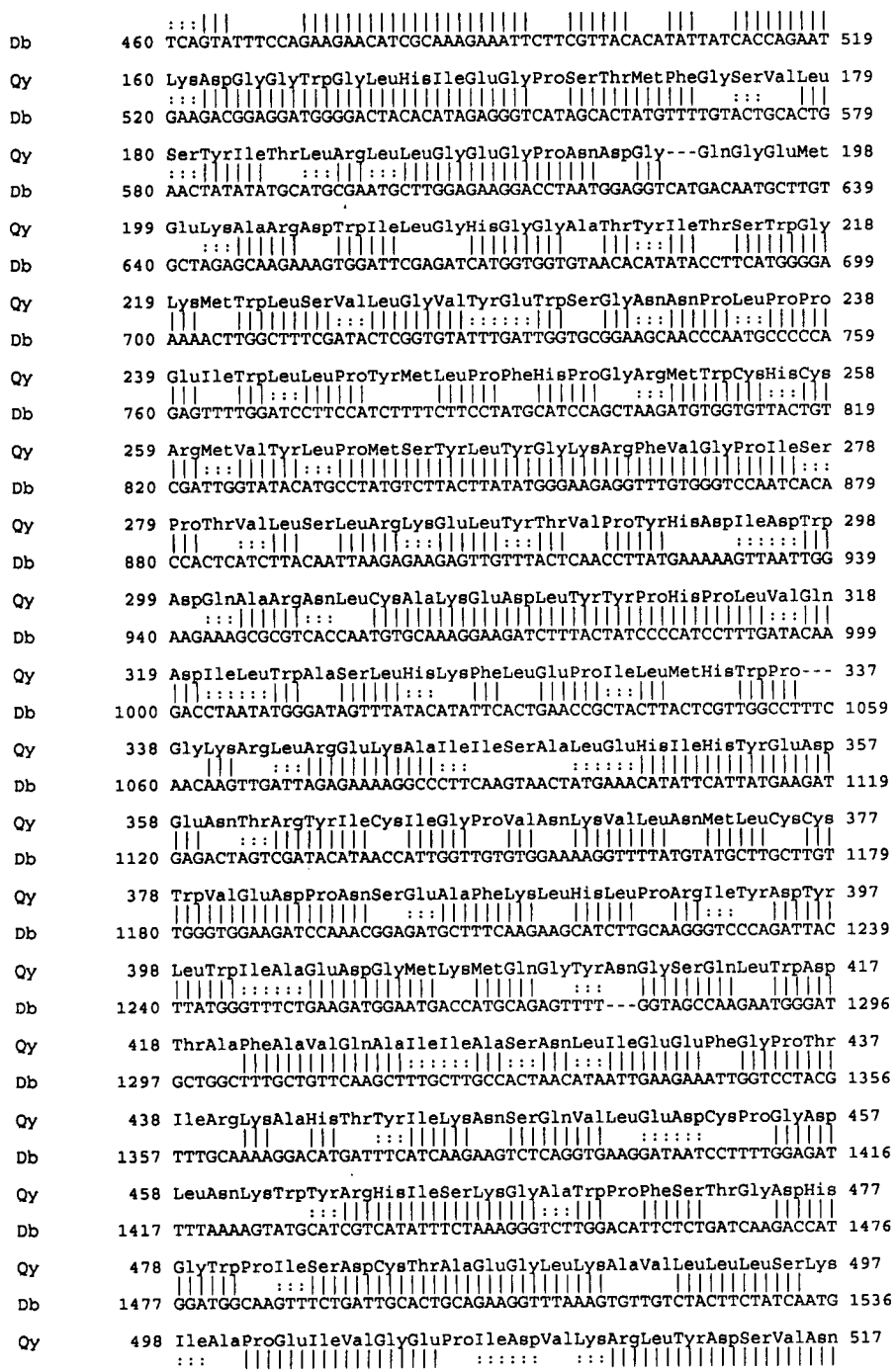
Search completed: November 5, 2005, 17:42:09
Job time : 1548 secs

Alignment Scores:

Pred. No.:	1.34e-284	Length:	2478
Score:	2538.00	Matches:	449
Percent Similarity:	75.79%	Conservative:	124
Best Local Similarity:	59.39%	Mismatches:	179
Query Match:	61.19%	Indels:	4
DB:	18	Gaps:	4

US-09-419-679-6 (1-757) x US-10-221-074-3 (1-2478)





RESULT 15
US-10-221-074-3
; Sequence 3, Application US/10221074
; Publication No. US20030208791A1
; GENERAL INFORMATION:
; APPLICANT: Bryan, Gregory T.
; APPLICANT: Hwang, Der-Chyan
; APPLICANT: Maxwell, Carl A.
; APPLICANT: McGonigle, Brian
; APPLICANT: Potter, Susan M.
; TITLE OF INVENTION: ENZYMES INVOLVED IN TRITERPENE SYNTHESIS
; FILE REFERENCE: BB1438 PCT
; CURRENT APPLICATION NUMBER: US/10/221,074
; CURRENT FILING DATE: 2003-01-08
; PRIOR APPLICATION NUMBER: 60/188,054
; PRIOR FILING DATE: 2000-03-09
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 3
; LENGTH: 2478
; TYPE: DNA
; ORGANISM: Glycine max
US-10-221-074-3

```
Qy        340 ArgLeuArgGluLysAlaIleIleSerAlaLeuGluHisIleHisTyrGluAspGluAsn 359
Db       1113 AAGCTGAGAGAGAGAGCTCTGGGAAACATCATGGAGCATATCCATTATGAAGATGAGAAC 1172

Qy        360 ThrArgTyrIleCysIleGlyProValAsnLysValLeuAsnMetLeuCysCysTrpVal 379
Db       1173 ACACAATACCTATGCATATGTCCTGTGAATAAGGCTCTAAACATGGTCTGCTGCTGGGTA 1232

Qy        380 GluAspProAsnSerGluAlaPheLysLeuHisLeuProArgIleTyrAspTyrLeuTrp 399
Db       1233 GAAGATCCAAATTCAGATGCATTCAAGCGTCACCTTGCAAGGATACCAGACTTCTTGTGG 1292

Qy        400 IleAlaGluAspGlyMetLysMetGlnGlyTyrAsnGlySerGlnLeuTrpAspThrAla 419
Db       1293 ATTTCAGAAGATGGCATGAAGGCACAAGTATATGATGGCTGCCAGAGCTGGGAGACATCA 1352

Qy        420 PheAlaValGlnAlaIleIleAlaSerAsnLeuIleGluGluPheGlyProThrIleArg 439
Db       1353 TTCATAATTCAAGCATTTTGCGCTACAGATCTTGTTAATGACTATGGTTCAACTCTTCAG 1412

Qy        440 LysAlaHisThrTyrIleLysAsnSerGlnValLeuGluAspCysProGlyAspLeuAsn 459
Db       1413 AGAGCCTATGAGTTTATGAAAAATTCACAGGTCATGAGGAACCATCCTGGTGACCAACGT 1472

Qy        460 LysTrpTyrArgHisIleSerLysGlyAlaTrpProPheSerThrGlyAspHisGlyTrp 479
Db       1473 TATTGGCATCGCCATAGATCGAAGGGTTCATGGACACTTTCATCTGCAGACAATGGATGG 1532

Qy        480 ProIleSerAspCysThrAlaGluGlyLeuLysAlaValLeuLeuLeuSerLysIleAla 499
Db       1533 GCTGTATCTGACACTACAGGAGAAGCACTTAAGGCTGTACTGTTGCTGTCAAAGATCTCA 1592

Qy        500 ProGlu------IleValGlyGluProIleAspValLysArgLeuTyrAspSerValAsn 517
Db       1593 AACAAAAACAACCTTGTTGGGGATCCAATAGAAAGAGAAAGGTTGCATGACGCTATTGAT 1652

Qy        518 ValIleLeuSerLeuGlnAsnGluAspGlyGlyPheAlaThrTyrGluLeuLysArgSer 537
Db       1653 TGCCTTCTATCTTTTGCGAACAAAGATGGCACCTTTTCTACATATGAGTGCAAAAGAACT 1712

Qy        538 TyrAsnTrpLeuGluIleIleAsnProAlaGluThrPheGlyAspIleValIleAspTyr 557
Db       1713 TATTCTTGGTTAGAGATTCTAAGTCCTTGTGAGACTTTCCCAAACATTGTTGTCGATTAT 1772

Qy        558 ProTyrValGluCysThrSerAlaAlaIleGlnAlaLeuAlaSerPheArgLysLeuTyr 577
Db       1773 CCTTACCCAGAATGCACTTCATCAGTGCTCCAAGCTCTGATATTGTTCAAAGACTTATAT 1832

Qy        578 ProGlyHisArgArgGluGluIleGlnHisCysIleAspLysAlaThrThrPheIleGlu 597
Db       1833 CCTGGTTACCGCACAGAAGAGATAGAAGCATTAGTTAGAAGTGCAGCAACGTTTATTGAG 1892

Qy        598 LysIleGlnAlaSerAspGlySerTrpTyrGlySerTrpGlyValCysPheThrTyrGly 617
Db       1893 ACCAAACAACAAGAAGACGGTTCATGGTTAGGTAATTGGGGTATATGTTTCACCTACGGG 1952

Qy        618 AlaTrpPheGlyValLysGlyLeuIleAlaAlaGlyArgSerPheSerAsnCysSerSer 637
Db       1953 GCCTTCTTTTCGATTAAAGGTTTAGTCGCTTCTGGAAGAACGTACAAGAACAGCCCTTGC 2012

Qy        638 IleArgLysAlaCysGluPheLeuLeuSerLysGlnLeuProSerGlyGlyTrpGlyGlu 657
Db       2013 ATACGGAAAGCATGCCACTTCATATTGTCGAAGCAGCTCAGTACTGGTGGATGGGGAGAA 2072

Qy        658 SerTyrLeuSerCysGlnAsnLysValTyrSerAsnLeuGluGlyAsnArgSerHisVal 677
Db       2073 AGCCATATCGCTATTGAAACTCAGGTGTATGTAAATCTCAAAGGTGACCGTGCTCACGCT 2132

Qy        678 ValAsnThrGlyTrpAlaMetLeuAlaLeuIleAspAlaGlyGlnAlaLysArgAspSer 697
Db       2133 GTGAATACCGCCTGGGCAATGCTCGCTTTAATTTATGCTGGACAGTTTGAACGAGATCCA 2192

Qy        698 GlnProLeuHisArgAlaAlaAlaTyrLeuIleAsnSerGlnLeuGluAspGlyAspPhe 717
Db       2193 ACGCCATTACATCGTGCTGCAAAGGAATTGATCAACATGCAACTGGAGACAGGAGAGTTT 2252

Qy        718 ProGlnGlnGluIleMetGlyValPheAsnLysAsnCysMetIleThrTyrAlaAlaTyr 737
Db       2253 CCTCAGCAAGAGCATGTAGGATGCTTCAACTGCAGCCTCTACTTCAATTACCCAAGCTAC 2312

Qy        738 ArgAsnIleProIleTrpAlaLeuGlyGluTyr 749
Db       2313 CGCAACTTGTTCCCCATCTGGGCTCTCGGTGAGTAC 2348

RESULT 14
US-10-424-599-35539
; Sequence 35539, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 35539
;   LENGTH: 2669
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_132095C.1
US-10-424-599-35539

Alignment Scores:
Pred. No.:                6.73e-285      Length:          2669
Score:                    2541.00        Matches:         450
Percent Similarity:       75.79%         Conservative:    123
Best Local Similarity:    59.52%         Mismatches:      179
Query Match:              61.26%         Indels:          4
DB:                       19             Gaps:            4

US-09-419-679-6 (1-757) x US-10-424-599-35539 (1-2669)

Qy          1 MetTrpLysLeuLysPheAlaGluGlyGlyAsn---ProTrpLeuArgThrLeuAsnAsn 19
Db         40 ATGTGGAGGCTGAAGATAGCAGATGGAGGAAATGATCCATACATATTCAGCACAAACAAT 99

Qy         20 HisValGlyArgGlnValTrpGluPheAspProLysLeuGlySerProGlnAspLeuLeu 39
Db        100 TTCGTTGGGAGGCAGACATGGGAGTTTGATCCTGAAGCAGGCAGTCCAGAGGAACGGGCC 159

Qy         40 GluIleGluLysAlaArgGlnAsnPheHisAspAsnArgPheThrHisLysHisSerAla 59
Db        160 CAGGTTGAAGCAGCTCGTCAGCATTTCTACCACAACCGCTTCAAGGTCAAGCCCTGCGCT 219

Qy         60 AspLeuLeuMetArgMetGlnPheAlaArgGluAsnProThrArgGluValLeuProLys 79
Db        220 GACCTCCTTTGGCGTTTTCAGGTTCTCAGAGAAAATAACTTCAAACAAACAATTCCTCGT 279

Qy         80 ValGlyValLysAspIleGluAspValThrGlnGluIleValThrLysThrLeuArgArg 99
Db        280 GTGACTATAGAAGATGGAGAGGAAATCACATACCAAAAAGTCACAAGCGCCGTCAGAAGG 339

Qy        100 AlaValSerPheHisSerThrLeuGlnCysHisAspGlyHisTrpProGlyAspTyrGly 119
Db        340 GGCGCACACCACCTTGCGGCACTGCAGACCTCTGATGGCCATTGGCCTGCTCAAATTGCA 399

Qy        120 GlyProMetPheLeuMetProGlyLeuValIleThrLeuSerIleThrGlyAlaLeuAsn 139
Db        400 GGTCCTCTCTTCTTTCTTCCTCCTTTGGTTTTTTGTATGTATATTACAGGAAATCTTGAA 459

Qy        140 ThrValLeuThrGluGluHisArgLysGluIleCysArgTyrLeuTyrAsnHisGlnAsn 159
```

```
Qy       538 TyrAsnTrpLeuGluIleIleAsnProAlaGluThrPheGlyAspIleValIleAspTyr 557
Db      1689 TATTCTTGGTTAGAGATTCTAAGTCCTTGTGAGACTTTCCCAAACATTGTTGTCGATTAT 1748

Qy       558 ProTyrValGluCysThrSerAlaAlaIleGlnAlaLeuAlaSerPheArgLysLeuTyr 577
Db      1749 CCTTACCCAGAATGCACTTCATCAGTGCTCCAAGCTCTGATATTGTTCAAAGACTTATAT 1808

Qy       578 ProGlyHisArgArgGluGluIleGlnHisCysIleAspLysAlaThrThrPheIleGlu 597
Db      1809 CCTGGTTACCGCACAGAAGAGATAGAAGCATTAGTTAGAAGTGCAGCAACGTTTATTGAG 1868

Qy       598 LysIleGlnAlaSerAspGlySerTrpTyrGlySerTrpGlyValCysPheThrTyrGly 617
Db      1869 ACCAAACAACAAGAAGACGGTTCATGGTTAGGTAATTGGGGTATATGTTTCACCTACGGG 1928

Qy       618 AlaTrpPheGlyValLysGlyLeuIleAlaAlaGlyArgSerPheSerAsnCysSerSer 637
Db      1929 GCCTTCTTTTCGATTAAAGGTTTAGTCGCTTCTGGAAGAACGTACAAGAACAGCCCTTGC 1988

Qy       638 IleArgLysAlaCysGluPheLeuLeuSerLysGlnLeuProSerGlyGlyTrpGlyGlu 657
Db      1989 ATACGGAAAGCATGCCACTTCATATTGTCGAAGCAGCTCAGTACTGGTGGATGGGGAGAA 2048

Qy       658 SerTyrLeuSerCysGlnAsnLysValTyrSerAsnLeuGluGlyAsnArgSerHisVal 677
Db      2049 AGCCATATCGCTATTGAAACTCAGGTGTATGTAAATCTCAAAGGTGACCGTGCTCACGCT 2108

Qy       678 ValAsnThrGlyTrpAlaMetLeuAlaLeuIleAspAlaGlyGlnAlaLysArgAspSer 697
Db      2109 GTGAATACCGCCTGGGCAATGCTCGCTTTAATTTATGCTGGACAGTTTGAACGAGATCCA 2168

Qy       698 GlnProLeuHisArgAlaAlaAlaTyrLeuIleAsnSerGlnLeuGluAspGlyAspPhe 717
Db      2169 ACGCCATTACATCGTGCTGCAAAGGAATTGATCAACATGCAACTGGAGACAGGAGAGTTT 2228

Qy       718 ProGlnGlnGluIleMetGlyValPheAsnLysAsnCysMetIleThrTyrAlaAlaTyr 737
Db      2229 CCTCAGCAAGAGCATGTAGGATGCTTCAACTGCAGCCTCTACTTCAATTACCCAAGCTAC 2288

Qy       738 ArgAsnIleProIleTrpAlaLeuGlyGluTyr 749
Db      2289 CGCAACTTGTTCCCCATCTGGGCTCTCGGTGAGTAC 2324
```

RESULT 13
US-10-425-115-65768
; Sequence 65768, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 65768
; LENGTH: 2627
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_159985C.1
US-10-425-115-65768

Alignment Scores:

Pred. No.:	3.35e-285	Length:	2627
Score:	2543.50	Matches:	444
Percent Similarity:	77.39%	Conservative:	138
Best Local Similarity:	59.04%	Mismatches:	167
Query Match:	61.32%	Indels:	3
DB:	21	Gaps:	2

US-09-419-679-6 (1-757) x US-10-425-115-65768 (1-2627)

```
Qy         1 MetTrpLysLeuLysPheAlaGluGlyGlyAsnProTrpLeuArgThrLeuAsnAsnHis 20
Db        93 ATGTGGAGGCTGACGGTGGCCGAGGGCGGCGGCCCGTGGCTGCGCTCGACGAACGGCTTC 152

Qy        21 ValGlyArgGlnValTrpGluPheAspProLysLeuGlySerProGlnAspLeuLeuGlu 40
Db       153 GTGGGGCGGGCGGTGTGGGAGTTCGACCCTGACCTCGGTACGCCGGAGGAGCGCGACGAG 212

Qy        41 IleGluLysAlaArgGlnAsnPheHisAspAsnArgPheThrHisLysHisSerAlaAsp 60
Db       213 GTAGAGAGGGTACGCCAGGAGTTCTCCGACCACCGCTTCCAGAGGAGAGAGTCGGCCGAC 272

Qy        61 LeuLeuMetArgMetGlnPheAlaArgGluAsnProThr---ArgGluValLeuProLys 79
Db       273 CTCCTCATGCGCATGCAGTGCGCAAAGCAGAACAGATCTCAACGTCGTGATCTGCCACGC 332

Qy        80 ValGlyValLysAspIleGluAspValThrGlnGluIleValThrLysThrLeuArgArg 99
Db       333 ATCAAGCTTGGGGAGGATGAGCACGTCACTGAAGAAATTGTACTGAGCTCCTTGAGGTCG 392

Qy       100 AlaValSerPheHisSerThrLeuGlnCysHisAspGlyHisTrpProGlyAspTyrGly 119
Db       393 GCTCTGGACCAGTTCTCTTCACTGCAAGCCAGCGATGGGCACTGGCCTGGTGATTTCAGC 452

Qy       120 GlyProMetPheLeuMetProGlyLeuValIleThrLeuSerIleThrGlyAlaLeuAsn 139
Db       453 GGGATTATGTTCATCATGCCTGGTTTGATATTTGCCTTGTATGTCACTGGATCACTGAAT 512

Qy       140 ThrValLeuThrGluGluHisArgLysGluIleCysArgTyrLeuTyrAsnHisGlnAsn 159
Db       513 GTTGTCATATCACCAGAACATCGGCATGAGATTTGCCGCTACATCTACAACCACCAGAAC 572

Qy       160 LysAspGlyGlyTrpGlyLeuHisIleGluGlyProSerThrMetPheGlySerValLeu 179
Db       573 GAAGATGGTGGATGGGGCACACTTATCTTGGGTTCGAGCACCATGTTTGGCACATGCTCA 632

Qy       180 SerTyrIleThrLeuArgLeuLeuGlyGluGlyProAsnAspGlyGlnGlyGluMetGlu 199
Db       633 AACTACATCACCTTGAGGCTTCTCGGTGAGGAGCCATACGCCAACAACAGTGTGTTGGCT 692

Qy       200 LysAlaArgAspTrpIleLeuGlyHisGlyGlyAlaThrTyrIleThrSerTrpGlyLys 219
Db       693 AAAGGGCGTGCTTGGATTCTATCCCATGGTGGTGCAACCTTGATTCCTCAGTGGGGAAAA 752

Qy       220 MetTrpLeuSerValLeuGlyValTyrGluTrpSerGlyAsnAsnProLeuProProGlu 239
Db       753 ATATGGCTTTCGGTACTTGGATTGTTTGATTGGTCAGGAAATAATCCAATTTTCCCTGAA 812

Qy       240 IleTrpLeuLeuProTyrMetLeuProPheHisProGlyArgMetTrpCysHisCysArg 259
Db       813 CTATGGTCCATTCCCCAGTTTCTTCCATTTCATCCAGGGAAATTCTGGTGCTTTGCCCGT 872

Qy       260 MetValTyrLeuProMetSerTyrLeuTyrGlyLysArgPheValGlyProIleSerPro 279
Db       873 ATGGTATATCTACCAATGGCTTATCTTTATGGCAAGAAATTTGTTGGACCAATTACACCA 932

Qy       280 ThrValLeuSerLeuArgLysGluLeuTyrThrValProTyrHisAspIleAspTrpAsp 299
Db       933 ACTATATTGGCACTAAGAGAGGAAATCTATGACACTCCTTATGGAAAGATTGACTGGAGT 992

Qy       300 GlnAlaArgAsnLeuCysAlaLysGluAspLeuTyrTyrProHisProLeuValGlnAsp 319
Db       993 GATGCTCGTAGTAAATGCGCAAAGGAGGACCTCATCTGTCCACGCACACTGTTGCAGAAT 1052

Qy       320 IleLeuTrpAlaSerLeuHisLysPheLeuGluProIleLeuMetHisTrpProGlyLys 339
Db      1053 GTTATTTGGACTTCACTTTATAGGTGTGTGGAACCAGTATTGAGCAGTTGGCCTATCAAC 1112
```

```
Qy        741 PheProIleTrpAlaLeuGlyGluTyrGlnSerGlnValLeu 754
              :::||||||||||||||||||||||:::::::   :::::::|||
Db       2271 TACCCCATTTGGGCTCTTGGGGAGTTTCGTCGTCGACTGCTT 2312

RESULT 12
US-10-425-114-24645
; Sequence 24645, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 24645
;   LENGTH: 2593
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: LIB3632-052-F1_FLI
US-10-425-114-24645

Alignment Scores:
Pred. No.:                3.28e-285      Length:          2593
Score:                    2543.50        Matches:         444
Percent Similarity:       77.39%         Conservative:    138
Best Local Similarity:    59.04%         Mismatches:      167
Query Match:              61.32%         Indels:          3
DB:                       19             Gaps:            2

US-09-419-679-6 (1-757) x US-10-425-114-24645 (1-2593)

Qy          1 MetTrpLysLeuLysPheAlaGluGlyGlyAsnProTrpLeuArgThrLeuAsnAsnHis 20
              ||||||:::|||     ||||||||||||   ||||||||||||:::   |||
Db         69 ATGTGGAGGCTGACGGTGGCCGAGGGCGGCGGCCCGTGGCTGCGCTCGACGAACGGCTTC 128

Qy         21 ValGlyArgGlnValTrpGluPheAspProLysLeuGlySerProGlnAspLeuLeuGlu 40
              |||||||||   ||||||||||||||||||   ||||||:::|||::::::      |||
Db        129 GTGGGGCGGGCGGTGTGGGAGTTCGACCCTGACCTCGGTACGCCGGAGGAGCGCGACGAG 188

Qy         41 IleGluLysAlaArgGlnAsnPheHisAspAsnArgPheThrHisLysHisSerAlaAsp 60
              :::|||:::   ||||||   |||   |||:::||||||      :::   |||||||||
Db        189 GTAGAGAGGGTACGCCAGGAGTTCTCCGACCACCGCTTCCAGAGGAGAGAGTCGGCCGAC 248

Qy         61 LeuLeuMetArgMetGlnPheAlaArgGluAsnProThr---ArgGluValLeuProLys 79
              ||||||||||||||||   |||:::::|||   :::   |||      ||||||:::
Db        249 CTCCTCATGCGCATGCAGTGCGCAAAGCAGAACAGATCTCAACGTCGTGATCTGCCACGC 308

Qy         80 ValGlyValLysAspIleGluAspValThrGlnGluIleValThrLysThrLeuArgArg 99
              :::   :::   :::   |||   ||||||:::||||||||      :::||||||
Db        309 ATCAAGCTTGGGGAGGATGAGCACGTCACTGAAGAAATTGTACTGAGCTCCTTGAGGTCG 368

Qy        100 AlaValSerPheHisSerThrLeuGlnCysHisAspGlyHisTrpProGlyAspTyrGly 119
              |||:::      |||:::||||||     ||||||||||||||||||:::
Db        369 GCTCTGGACCAGTTCTCTTCACTGCAAGCCAGCGATGGGCACTGGCCTGGTGATTTCAGC 428

Qy        120 GlyProMetPheLeuMetProGlyLeuValIleThrLeuSerIleThrGlyAlaLeuAsn 139
              |||   ||||||:::||||||||||:::      |||   :::||||||:::||||||
Db        429 GGGATTATGTTCATCATGCCTGGTTTGATATTTGCCTTGTATGTCACTGGATCACTGAAT 488

Qy        140 ThrValLeuThrGluGluHisArgLysGluIleCysArgTyrLeuTyrAsnHisGlnAsn 159
                 |||::::::   |||||||||   ||||||||||||||:::||||||||||||||||
Db        489 GTTGTCATATCACCAGAACATCGGCATGAGATTTGCCGCTACATCTACAACCACCAGAAC 548

Qy        160 LysAspGlyGlyTrpGlyLeuHisIleGluGlyProSerThrMetPheGlySerValLeu 179
              :::||||||||||||||||      |||   |||   ||||||||||||||||:::

Db        549 GAAGATGGTGGATGGGGCACACTTATCTTGGGTTCGAGCACCATGTTTGGCACATGCTCA 608

Qy        180 SerTyrIleThrLeuArgLeuLeuGlyGluGlyProAsnAspGlyGlnGlyGluMetGlu 199
              :::||||||||||||||||||||||||||||||||   |||                :::
Db        609 AACTACATCACCTTGAGGCTTCTCGGTGAGGAGCCATACGCCAACAACAGTGTGTTGGCT 668

Qy        200 LysAlaArgAspTrpIleLeuGlyHisGlyGlyAlaThrTyrIleThrSerTrpGlyLys 219
              |||   |||   |||||||||   ||||||||||||||||   |||      ||||||||
Db        669 AAAGGGCGTGCTTGGATTCTATCCCATGGTGGTGCAACCTTGATTCCTCAGTGGGGAAAA 728

Qy        220 MetTrpLeuSerValLeuGlyValTyrGluTrpSerGlyAsnAsnProLeuProProGlu 239
              :::||||||||||||||||||:::::::::||||||||||||||||||||:::   ||||||
Db        729 ATATGGCTTTCGGTACTTGGATTGTTTGATTGGTCAGGAAATAATCCAATTTTCCCTGAA 788

Qy        240 IleTrpLeuLeuProTyrMetLeuProPheHisProGlyArgMetTrpCysHisCysArg 259
              :::|||   :::|||      ||||||||||||||||||:::   ||||||      |||
Db        789 CTATGGTCCATTCCCCAGTTTCTTCCATTTCATCCAGGGAAATTCTGGTGCTTTGCCCGT 848

Qy        260 MetValTyrLeuProMetSerTyrLeuTyrGlyLysArgPheValGlyProIleSerPro 279
              ||||||||||||||||||:::||||||||||||||:::||||||||||||||||:::|||
Db        849 ATGGTATATCTACCAATGGCTTATCTTTATGGCAAGAAATTTGTTGGACCAATTACACCA 908

Qy        280 ThrValLeuSerLeuArgLysGluLeuTyrThrValProTyrHisAspIleAspTrpAsp 299
              |||:::|||:::||||||:::|||:::|||      ||||||      |||||||||
Db        909 ACTATATTGGCACTAAGAGAGGAAATCTATGACACTCCTTATGGAAAGATTGACTGGAGT 968

Qy        300 GlnAlaArgAsnLeuCysAlaLysGluAspLeuTyrTyrProHisProLeuValGlnAsp 319
              ||||||:::   ||||||||||||||||||      |||      |||:::|||:::
Db        969 GATGCTCGTAGTAAATGCGCAAAGGAGGACCTCATCTGTCCACGCACACTGTTGCAGAAT 1028

Qy        320 IleLeuTrpAlaSerLeuHisLysPheLeuGluProIleLeuMetHisTrpProGlyLys 339
              ::::::|||   ||||||::::::   :::|||||||:::|||      ||||||
Db       1029 GTTATTTGGACTTCACTTTATAGGTGTGTGGAACCAGTATTGAGCAGTTGGCCTATCAAC 1088

Qy        340 ArgLeuArgGluLysAlaIleIleSerAlaLeuGluHisIleHisTyrGluAspGluAsn 359
              :::||||||||:::|||:::   :::   :::|||||||||||||||||||||||||||
Db       1089 AAGCTGAGAGAGAGAGCTCTGGGAAACATCATGGAGCATATCCATTATGAAGATGAGAAC 1148

Qy        360 ThrArgTyrIleCysIleGlyProValAsnLysValLeuAsnMetLeuCysCysTrpVal 379
              |||:::|||:::|||||||   |||||||||||   |||||||||:::||||||||||
Db       1149 ACACAATACCTATGCATATGTCCTGTGAATAAGGCTCTAAACATGGTCTGCTGCTGGGTA 1208

Qy        380 GluAspProAsnSerGluAlaPheLysLeuHisLeuProArgIleTyrAspTyrLeuTrp 399
              ||||||||||||||:::|||||||||   ||||||   ||||||   |||:::||||||
Db       1209 GAAGATCCAAATTCAGATGCATTCAAGCGTCACCTTGCAAGGATACCAGACTTCTTGTGG 1268

Qy        400 IleAlaGluAspGlyMetLysMetGlnGlyTyrAsnGlySerGlnLeuTrpAspThrAla 419
              |||:::|||||||||||||||   |||   |||:::|||   |||   |||:::|||:::
Db       1269 ATTTCAGAAGATGGCATGAAGGCACAAGTATATGATGGCTGCCAGAGCTGGGAGACATCA 1328

Qy        420 PheAlaValGlnAlaIleIleAlaSerAsnLeuIleGluGluPheGlyProThrIleArg 439
              |||   :::||||||      |||::::::|||:::   ::::::|||   |||::::::
Db       1329 TTCATAATTCAAGCATTTTGCGCTACAGATCTTGTTAATGACTATGGTTCAACTCTTCAG 1388

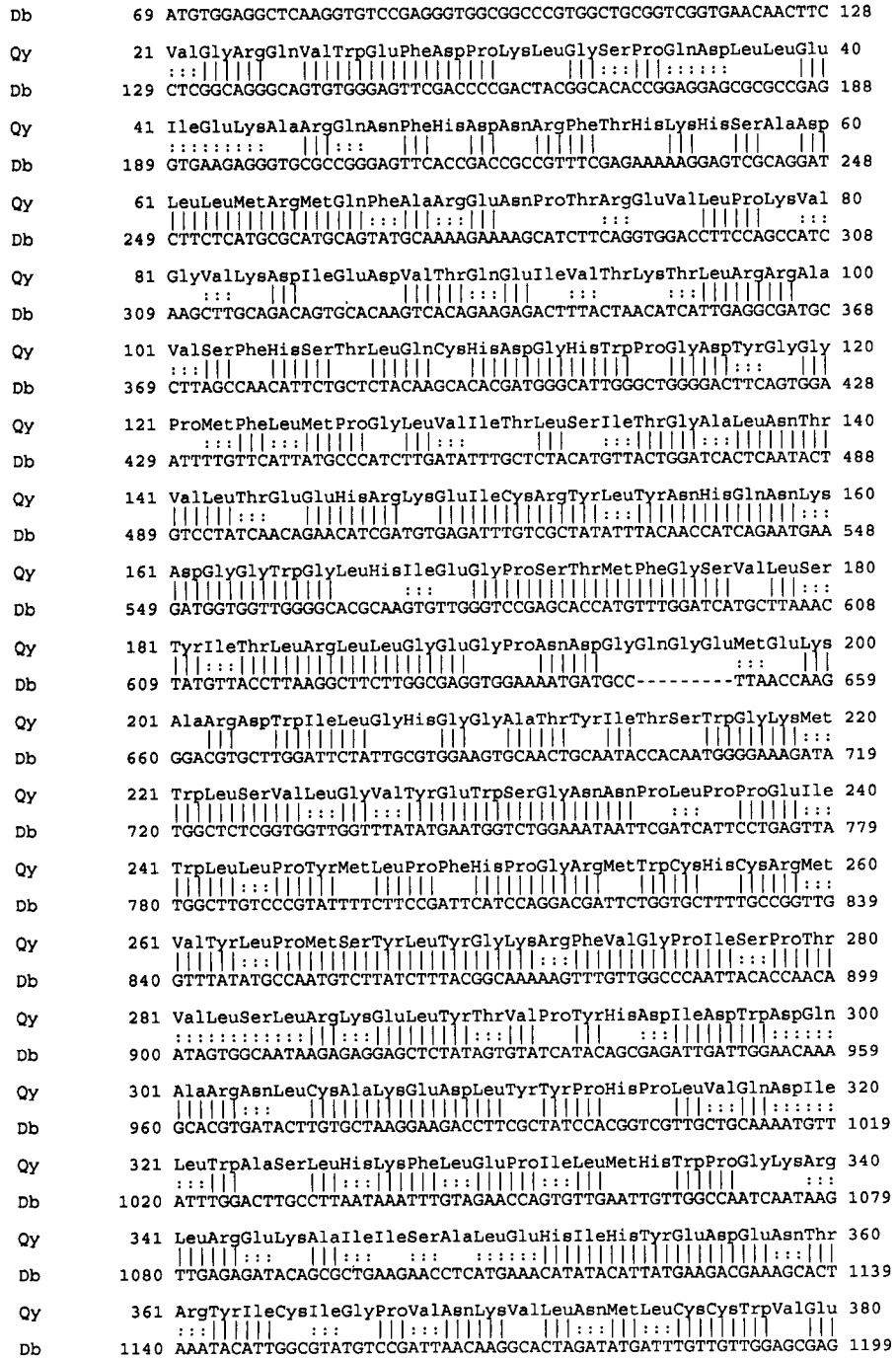
Qy        440 LysAlaHisThrTyrIleLysAsnSerGlnValLeuGluAspCysProGlyAspLeuAsn 459
              :::|||:::   ::::::|||||||||||||||:::   :::   |||||||||
Db       1389 AGAGCCTATGAGTTTATGAAAAATTCACAGGTCATGAGGAACCATCCTGGTGACCAACGT 1448

Qy        460 LysTrpTyrArgHisIleSerLysGlyAlaTrpProPheSerThrGlyAspHisGlyTrp 479
              |||:::||||||   ||||||||||:::|||      |||:::   |||:::||||||
Db       1449 TATTGGCATCGCCATAGATCGAAGGGTTCATGGACACTTTCATCTGCAGACAATGGATGG 1508

Qy        480 ProIleSerAspCysThrAlaGluGlyLeuLysAlaValLeuLeuLeuSerLysIleAla 499
              :::||||||   |||   |||||||||||||||||||||||||||||||||||||:::
Db       1509 GCTGTATCTGACACTACAGGAGAAGCACTTAAGGCTGTACTGTTGCTGTCAAAGATCTCA 1568

Qy        500 ProGlu------IleValGlyGluProIleAspValLysArgLeuTyrAspSerValAsn 517
              :::      :::||||||:::||||||:::   :::||||||:::|||:::::::::
Db       1569 AACAAAAACAACCTTGTTGGGGATCCAATAGAAAGAGAAAGGTTGCATGACGCTATTGAT 1628

Qy        518 ValIleLeuSerLeuGlnAsnGluAspGlyGlyPheAlaThrTyrGluLeuLysArgSer 537
              :::||||||      |||:::||||||   |||:::||||||||||   ||||||:::
Db       1629 TGCCTTCTATCTTTTGCGAACAAAGATGGCACCTTTTCTACATATGAGTGCAAAAGAACT 1688
```


Db 69 ATGTGGAGGCTCAAGGTGTCCGAGGGTGGCGGCCCGTGGCTGCGGTCGGTGAACAACTTC 128
Qy 21 ValGlyArgGlnValTrpGluPheAspProLysLeuGlySerProGlnAspLeuLeuGlu 40
Db 129 CTCGGCAGGGCAGTGTGGGAGTTCGACCCCGACTACGGCACACCGGAGGAGCGCGCCGAG 188
Qy 41 IleGluLysAlaArgGlnAsnPheHisAspAsnArgPheThrHisLysHisSerAlaAsp 60
Db 189 GTGAAGAGGGTGCGCCGGGAGTTCACCGACCGCCGTTTCGAGAAAAAGGAGTCGCAGGAT 248
Qy 61 LeuLeuMetArgMetGlnPheAlaArgGluAsnProThrArgGluValLeuProLysVal 80
Db 249 CTTCTCATGCGCATGCAGTATGCAAAAGAAAAGCATCTTCAGGTGGACCTTCCAGCCATC 308
Qy 81 GlyValLysAspIleGluAspValThrGlnGluIleValThrLysThrLeuArgArgAla 100
Db 309 AAGCTTGCAGACAGTGCACAAGTCACAGAAGAGACTTTACTAACATCATTGAGGCGATGC 368
Qy 101 ValSerPheHisSerThrLeuGlnCysHisAspGlyHisTrpProGlyAspTyrGlyGly 120
Db 369 CTTAGCCAACATTCTGCTCTACAAGCACACGATGGGCATTGGGCTGGGGACTTCAGTGGA 428
Qy 121 ProMetPheLeuMetProGlyLeuValIleThrLeuSerIleThrGlyAlaLeuAsnThr 140
Db 429 ATTTTGTTCATTATGCCCATCTTGATATTTGCTCTACATGTTACTGGATCACTCAATACT 488
Qy 141 ValLeuThrGluGluHisArgLysGluIleCysArgTyrLeuTyrAsnHisGlnAsnLys 160
Db 489 GTCCTATCAACAGAACATCGATGTGAGATTTGTCGCTATATTTACAACCATCAGAATGAA 548
Qy 161 AspGlyGlyTrpGlyLeuHisIleGluGlyProSerThrMetPheGlySerValLeuSer 180
Db 549 GATGGTGGTTGGGGCACGCAAGTGTTGGGTCCGAGCACCATGTTTGGATCATGCTTAAAC 608
Qy 181 TyrIleThrLeuArgLeuLeuGlyGluGlyProAsnAspGlyGlnGlyGluMetGluLys 200
Db 609 TATGTTACCTTAAGGCTTCTTGGCGAGGTGGAAAATGATGCC---------TTAACCAAG 659
Qy 201 AlaArgAspTrpIleLeuGlyHisGlyGlyAlaThrTyrIleThrSerTrpGlyLysMet 220
Db 660 GGACGTGCTTGGATTCTATTGCGTGGAAGTGCAACTGCAATACCACAATGGGGAAAGATA 719
Qy 221 TrpLeuSerValLeuGlyValTyrGluTrpSerGlyAsnAsnProLeuProProGluIle 240
Db 720 TGGCTCTCGGTGGTTGGTTTATATGAATGGTCTGGAAATAATTCGATCATTCCTGAGTTA 779
Qy 241 TrpLeuLeuProTyrMetLeuProPheHisProGlyArgMetTrpCysHisCysArgMet 260
Db 780 TGGCTTGTCCCGTATTTTCTTCCGATTCATCCAGGACGATTCTGGTGCTTTTGCCGGTTG 839
Qy 261 ValTyrLeuProMetSerTyrLeuTyrGlyLysArgPheValGlyProIleSerProThr 280
Db 840 GTTTATATGCCAATGTCTTATCTTTACGGCAAAAAGTTTGTTGGCCCAATTACACCAACA 899
Qy 281 ValLeuSerLeuArgLysGluLeuTyrThrValProTyrHisAspIleAspTrpAspGln 300
Db 900 ATAGTGGCAATAAGAGAGGAGCTCTATAGTGTATCATACAGCGAGATTGATTGGAACAAA 959
Qy 301 AlaArgAsnLeuCysAlaLysGluAspLeuTyrTyrProHisProLeuValGlnAspIle 320
Db 960 GCACGTGATACTTGTGCTAAGGAAGACCTTCGCTATCCACGGTCGTTGCTGCAAAATGTT 1019
Qy 321 LeuTrpAlaSerLeuHisLysPheLeuGluProIleLeuMetHisTrpProGlyLysArg 340
Db 1020 ATTTGGACTTGCCTTAATAAATTTGTAGAACCAGTGTTGAATTGTTGGCCAATCAATAAG 1079
Qy 341 LeuArgGluLysAlaIleIleSerAlaLeuGluHisIleHisTyrGluAspGluAsnThr 360
Db 1080 TTGAGAGATACAGCGCTGAAGAACCTCATGAAACATATACATTATGAAGACGAAAGCACT 1139
Qy 361 ArgTyrIleCysIleGlyProValAsnLysValLeuAsnMetLeuCysCysTrpValGlu 380
Db 1140 AAATACATTGGCGTATGTCCGATTAACAAGGCACTAGATATGATTTGTTGTTGGAGCGAG 1199
Qy 381 AspProAsnSerGluAlaPheLysLeuHisLeuProArgIleTyrAspTyrLeuTrpIle 400
Db 1200 GATCCAAATTCAGATGCACTGAAGTTGCATCTTCCAAGGATCTATGACTATTTATGGCTT 1259
Qy 401 AlaGluAspGlyMetLysMetGlnGlyTyrAsnGlySerGlnLeuTrpAspThrAlaPhe 420
Db 1260 GCAGAAGATGGCATGAAAGCACAGGTTTATGATGGTTGTCAAAGCTGGGAGCTTGCTTTT 1319
Qy 421 AlaValGlnAlaIleIleAlaSerAsnLeuIleGluGluPheGlyProThrIleArgLys 440
Db 1320 ATTGTTCAAGCATATTGCTCGACAGACCTTGTTAATGAGTTTGGTCCAACACTTCGGAAA 1379
Qy 441 AlaHisThrTyrIleLysAsnSerGlnValLeuGluAspCysProGlyAspLeuAsnLys 460
Db 1380 GCCCATGAGTTCATTAAAAGTTCACAGGTTCTTGAGAACCATCCTAACAGT---GAAACT 1436
Qy 461 TrpTyrArgHisIleSerLysGlyAlaTrpProPheSerThrGlyAspHisGlyTrpPro 480
Db 1437 TATTACCGCCATAGGTCAAAAGGTTCATGGACACTTTCAACAGCGGATAATGGGTGGTCT 1496
Qy 481 IleSerAspCysThrAlaGluGlyLeuLysAlaValLeuLeuLeuSerLysIleAlaPro 500
Db 1497 GTATCAGATTGTACTGCAGAAGCACTTAAGGCATTGTTGTTGTTGTCGAAGATCTCTCCT 1556
Qy 501 GluIleValGlyGluProIleAspValLysArgLeuTyrAspSerValAsnValIleLeu 520
Db 1557 AATCTTGTGGGGGATCCCGTAAAAGGAGAAAGGTTGCATGATGCAGTCGATTGCTTACTT 1616
Qy 521 SerLeuGlnAsnGluAspGlyGlyPheAlaThrTyrGluLeuLysArgSerTyrAsnTrp 540
Db 1617 TCTTTTATGAATAAAGATGGCACATTTTCTACATATGAGTGTAAGAGAACTACATCTCTA 1676
Qy 541 LeuGluIleIleAsnProAlaGluThrPheGlyAspIleValIleAspTyrProTyrVal 560
Db 1677 TTAGAGGTTCTCAACCCTTCTGAAAGTTTTCTGAACATTATTGTCGACTATCCATCTGTC 1736
Qy 561 GluCysThrSerAlaAlaIleGlnAlaLeuAlaSerPheArgLysLeuTyrProGlyHis 580
Db 1737 GAATGTACATCATCTGTGCTTCAGGCCCTAATTATGTTCAAAGAGCTTTACCCTGGGTAC 1796
Qy 581 ArgArgGluGluIleGlnHisCysIleAspLysAlaThrThrPheIleGluLysIleGln 600
Db 1797 CGCAAAGAAGAGATAGGAAAATGTATTAAAAATGCTTCCAAGTTCATTGAGGACAAGCAA 1856
Qy 601 AlaSerAspGlySerTrpTyrGlySerTrpGlyValCysPheThrTyrGlyAlaTrpPhe 620
Db 1857 CGAAAGGATGGCTCATGGTTTGGCACTTGGGGTATATGTTTCACTTATGGGACGTTCTTT 1916
Qy 621 GlyValLysGlyLeuIleAlaAlaGlyArgSerPheSerAsnCysSerSerIleArgLys 640
Db 1917 GGTGTAAAAGGATTGATTGCTTCTGGAAGAACTTACGAAAATAGTTCTTCCATAAGGAAA 1976
Qy 641 AlaCysGluPheLeuLeuSerLysGlnLeuProSerGlyGlyTrpGlyGluSerTyrLeu 660
Db 1977 GCATGCAATTTTCTGTTGTCAAAGCAACTAAGTACAGGTGGATGGGGAGAGTCTTATCTT 2036
Qy 661 SerCysGlnAsnLysValTyrSerAsnLeuGluGlyAsnArgSerHisValValAsnThr 680
Db 2037 TCTAGTGAAACTGAGGCTTAT------GTGGAGGCCACTAGTCCTCATGCAGTGAACACT 2090
Qy 681 GlyTrpAlaMetLeuAlaLeuIleAspAlaGlyGlnAlaLysArgAspSerGlnProLeu 700
Db 2091 GCTTGGGCAATGTTGGCTTTAATTTATGCTGGGCAGGTTGAACGAGATCCTACTCCACTA 2150
Qy 701 HisArgAlaAlaAlaTyrLeuIleAsnSerGlnLeuGluAspGlyAspProGlnGln 720
Db 2151 TATCATGCTGCAAAAGAGTTGATCAATATGCAACTAGAGACAGGAGAGTTTCCCCAGCAA 2210
Qy 721 GluIleMetGlyValPheAsnLysAsnCysMetIleThrTyrAlaAlaTyrArgAsnIle 740
Db 2211 GAACACGTTGGATGCTTCAACTGCTCCATATACTTTAATTACGGCAACTATCGCAACTTA 2270

```
Db        483 ATAACATCGTGGGGGAAGTTTTGGCTTTCGGTACTAGGTGTATTTGAATGGTCTGGTAAC 542

Qy        234 AsnProLeuProProGluIleTrpLeuLeuProTyrMetLeuProPheHisProGlyArg 253

Db        543 AACCCGGTGCCACCAGAAGTATGGCTACTGCCATATCTCCTCCCATTTCACCCAGGGAGG 602

Qy        254 MetTrpCysHisCysArgMetValTyrLeuProMetSerTyrLeuTyrGlyLysArgPhe 273

Db        603 ATGTGGTGTCATTGTCGAATGGTGTATTTGCCAATGTGCTACATTTATGGGAAGAGGTTT 662

Qy        274 ValGlyProIleSerProThrValLeuSerLeuArgLysGluLeuTyrThrValProTyr 293

Db        663 GTTGGCCGAATCACACCACTTCTGTTGGAATTAAGGAAGGAACTTTTCAAAGACCCTTAC 722

Qy        294 HisAspIleAspTrpAspGlnAlaArgAsnLeuCysAlaLysGluAspLeuTyrTyrPro 313

Db        723 AGCAAGATTGATTGGGACAAGGCCCGCAACCTATGTGCCAAGGAAGATCTGTACTACCCA 782

Qy        314 HisProLeuValGlnAspIleLeuTrpAlaSerLeuHisLysPheLeuGluProIleLeu 333

Db        783 CACCCATTCGTTCAAGATGTGTTGTGGGCCACTCTCCATAAATTCGTTGAACCAGTTATG 842

Qy        334 MetHisTrpProGlyLysArgLeuArgGluLysAlaIleIleSerAlaLeuGluHisIle 353

Db        843 ATGCATTGGCCTGGCAGCAAATTGAGGGAGAAAGCTCTGGAAACAGTCATGCAACATGTT 902

Qy        354 HisTyrGluAspGluAsnThrArgTyrIleCysIleGlyProValAsnLysValLeuAsn 373

Db        903 CATTATGAAGATGAGAACACTCGTTATATTTGCATTGGTCCTGTAAACAAGGTATTGAAT 962

Qy        374 MetLeuCysCysTrpValGluAspProAsnSerGluAlaPheLysLeuHisLeuProArg 393

Db        963 ATGCTTGCTTGCTGGATTGAAGATCCAAACTCGGAGGCCTTCAAACTTCATATCCCACGA 1022

Qy        394 IleTyrAspTyrLeuTrpIleAlaGluAspGlyMetLysMetGlnGlyTyrAsnGlySer 413

Db       1023 GTCTATGATTACTTGTGGCTTGCTGAAGATGGCATGAAGATGCAGGGTTATAATGGCAGC 1082

Qy        414 GlnLeuTrpAspThrAlaPheAlaValGlnAlaIleIleAlaSerAsnLeuIleGluGlu 433

Db       1083 CAACTTTGGGATACAGCTTTCACAGTTCAAGCCATTGTGGCTACCAACCTTATTGAAGAG 1142

Qy        434 PheGlyProThrIleArgLysAlaHisThrTyrIleLysAsnSerGlnValLeuGluAsp 453

Db       1143 TTTGGTCCTACCCTTAAACTAGCACACAACTATATCAAGAATTCACAGGTTCTTGATGAC 1202

Qy        454 CysProGlyAspLeuAsnLysTrpTyrArgHisIleSerLysGlyAlaTrpProPheSer 473

Db       1203 TGCCCTGGTGATCTGAATGACTGGTACCGCCACACATCTAAAGGTGCATGGCCATTCTCA 1262

Qy        474 ThrGlyAspHisGlyTrpProIleSerAspCysThrAlaGluGlyLeuLysAlaValLeu 493

Db       1263 ACTGCTGATCATGGTTGGCCTATATCTGATTGCACTGCTGAAGGACTAAAGGCTTCATTA 1322

Qy        494 LeuLeuSerLysIleAlaProGluIleValGlyGluProIleAspValLysArgLeuTyr 513

Db       1323 TTGTTATCAAGGATCTCTCCCAAAATTGTTGGTGAACCGATGGAAGCTAATAGATTTTAT 1382

Qy        514 AspSerValAsnValIleLeuSerLeuGlnAsnGluAspGlyGlyPheAlaThrTyrGlu 533

Db       1383 GATGCTGTCAGTTGTCTGATGTCTTATATGAATGATAATGGCGGTTTCGCGACATATGAA 1442

Qy        534 LeuLysArgSerTyrAsnTrpLeuGluIleIleAsnProAlaGluThrPheGlyAspIle 553

Db       1443 CTCACAAGATCTTATCCCTGGTTGGAGCTGATCAATCCCGCTGAGACCTTTGGGGATATT 1502

Qy        554 ValIleAspTyrProTyrValGluCysThrSerAlaAlaIleGlnAlaLeuAlaSerPhe 573

Db       1503 GTGATTGATTACCCGTATGTTGAATGTACATCAGCAGCAATTCAGGCCCTGACATCATTC 1562

Qy        574 ArgLysLeuTyrProGlyHisArgArgGluGluIleGlnHisCysIleAspLysAlaThr 593

Db       1563 AAAAAACTATACCCTGGGCACCGCAGGAAAGAGGTGGATAACTGTATCAGCAAAGCTTCC 1622

Qy        594 ThrPheIleGluLysIleGlnAlaSerAspGlySerTrpTyrGlySerTrpGlyValCys 613

Db       1623 AATTTCATCGAGAGTATTCAGAAAAGCGATGGTTCATGGTATGGCTCTTGGGCCGTCTGT 1682

Qy        614 PheThrTyrGlyAlaTrpPheGlyValLysGlyLeuIleAlaAlaGlyArgSerPheSer 633

Db       1683 TTCACATACGGCACTTGGTTTGGTGTGAAGGGACTAATTGCTGCTGGTAGAACATTTGAG 1742

Qy        634 AsnCysSerSerIleArgLysAlaCysGluPheLeuLeuSerLysGlnLeuProSerGly 653

Db       1743 AACAGTCCTGCAATTAGAAAGGCATGCGACTTTCTGTTGTCAAAAGAACTTCCTTCCGGT 1802

Qy        654 GlyTrpGlyGluSerTyrLeuSerCysGlnAsnLysValTyrSerAsnLeuGluGlyAsn 673

Db       1803 GGTTGGGGAGAAAGCTATTTGTCATCTCAAGACCAGGTTTACACCAATCTCAAAGGCAAC 1862

Qy        674 ArgSerHisValValAsnThrGlyTrpAlaMetLeuAlaLeuIleAspAlaGlyGlnAla 693

Db       1863 CGGCCTCATGCGGTGAACACTAGTTGGGCCATGCTGGCGCTGATTGATGCGGGCCAGGCC 1922

Qy        694 LysArgAspSerGlnProLeuHisArgAlaAlaAlaTyrLeuIleAsnSerGlnLeuGlu 713

Db       1923 GAGAGAGATCCAACGCCTCTACACCGAGCAGCAAAGGTTTTGATCAACTTACAATCAGAG 1982

Qy        714 AspGlyAspProProGlnGlnGluIleMetGlyValPheAsnLysAsnCysMetIleThr 733

Db       1983 GACGGAGAATTTCCTCAGCAAGAGATCATAGGAGTGTTCAACAAGAACTGCATGATAAGC 2042

Qy        734 TyrAlaAlaTyrArgAsnIleProIleTrpAlaLeuGlyGluTyrGlnSerGlnVal 753

Db       2043 TACTCCCAGTACAGGAACATCTTCCCGATTTGGGCTCTGGGTGAGTACCGGTGTCGAGTC 2102

Qy        754 Leu 754

Db       2103 TTG 2105
```

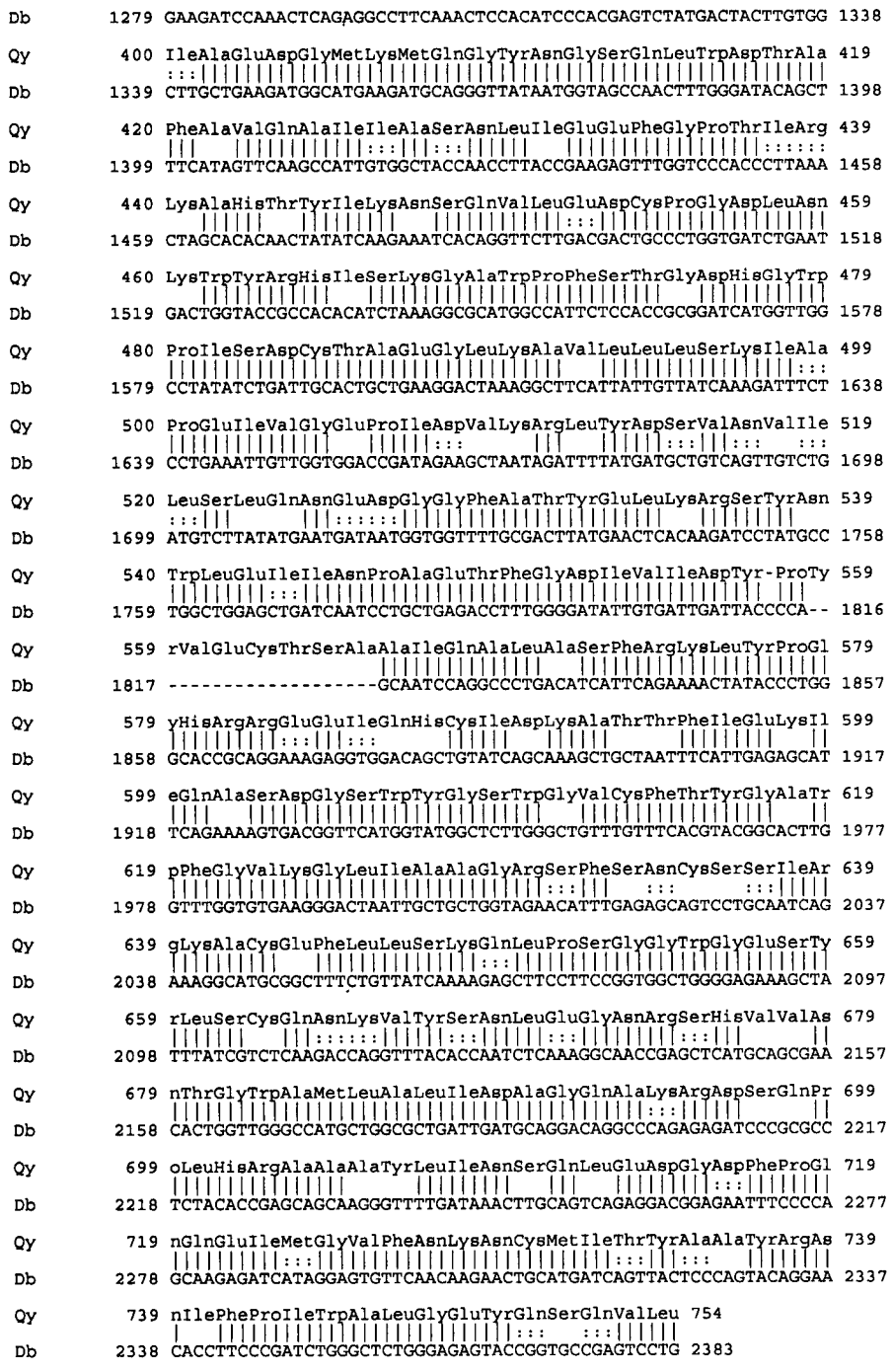
RESULT 11
US-10-221-074-7
; Sequence 7, Application US/10221074
; Publication No. US20030208791A1
; GENERAL INFORMATION:
; APPLICANT: Bryan, Gregory T.
; APPLICANT: Hwang, Der-Chyan
; APPLICANT: Maxwell, Carl A.
; APPLICANT: McGonigle, Brian
; APPLICANT: Potter, Susan M.
; TITLE OF INVENTION: ENZYMES INVOLVED IN TRITERPENE SYNTHESIS
; FILE REFERENCE: BB1438 PCT
; CURRENT APPLICATION NUMBER: US/10/221,074
; CURRENT FILING DATE: 2003-01-08
; PRIOR APPLICATION NUMBER: 60/188,054
; PRIOR FILING DATE: 2000-03-09
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 7
; LENGTH: 2538
; TYPE: DNA
; ORGANISM: Triticum aestivum
US-10-221-074-7

Alignment Scores:

Pred. No.:	4.23e-286	Length:	2538
Score:	2551.00	Matches:	446
Percent Similarity:	77.59%	Conservative:	139
Best Local Similarity:	59.15%	Mismatches:	163
Query Match:	61.50%	Indels:	6
DB:	18	Gaps:	3

US-09-419-679-6 (1-757) x US-10-221-074-7 (1-2538)

```
Qy          1 MetTrpLysLeuLysPheAlaGluGlyGlyAsnProTrpLeuArgThrLeuAsnAsnHis 20
```

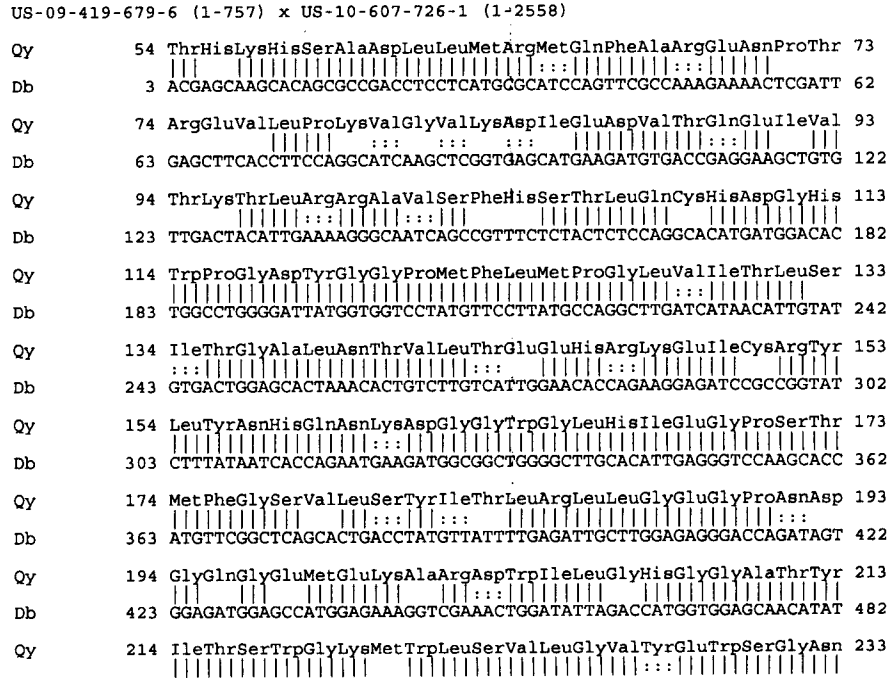


RESULT 10
US-10-607-726-1
; Sequence 1, Application US/10607726
; Publication No. US20040121439A1
; GENERAL INFORMATION:
; APPLICANT: Falco, S. Carl
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Hitz, William D.
; APPLICANT: Kinney, Anthony J.
; APPLICANT: Rafalski, Antoni
; APPLICANT: McGonigle, Brian
; APPLICANT: Lohman, Karin
; TITLE OF INVENTION: Enzymes Involved In Squalene Metabolizm
; FILE REFERENCE: BB1113 US NA
; CURRENT APPLICATION NUMBER: US/10/607,726
; CURRENT FILING DATE: 2003-06-27
; PRIOR APPLICATION NUMBER: US/09/419,679
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/105,405
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-10-23
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 1
; LENGTH: 2558
; TYPE: DNA
; ORGANISM: Zea mays
US-10-607-726-1

Alignment Scores:

Pred. No.:	0	Length:	2558
Score:	3125.00	Matches:	545
Percent Similarity:	88.87%	Conservative:	78
Best Local Similarity:	77.75%	Mismatches:	78
Query Match:	75.34%	Indels:	0
DB:	20	Gaps:	0

US-09-419-679-6 (1-757) x US-10-607-726-1 (1-2558)



```
Qy        595 PheIleGluLysIleGlnAlaSerAspGlySerTrpTyrGlySerTrpGlyValCysPhe 614
Db       1861 TTTATTGAGGGTATTCAAAAAAGCGATGGTTCATGGTATGGTTCTTGGGCTGTTTGTTTT 1920

Qy        615 ThrTyrGlyAlaTrpPheGlyValLysGlyLeuIleAlaAlaGlyArgSerPheSerAsn 634
Db       1921 ACCTATGGCACATGGTTTGGTGTAAAGGGATTAGTTGCTGCTGGTAGGACATTCAAAAAC 1980

Qy        635 CysSerSerIleArgLysAlaCysGluPheLeuLeuSerLysGlnLeuProSerGlyGly 654
Db       1981 AGTCCTGCAATCAGAAAGGCATGTGACTTTTTGTTGTCAAAAGAGCTTCCTTCTGGAGGC 2040

Qy        655 TrpGlyGluSerTyrLeuSerCysGlnAsnLysValTyrSerAsnLeuGluGlyAsnArg 674
Db       2041 TGGGGAGAAAGCTATTTGTCATCCCAAGATCAGGTTTATACCAATCTCGAAGGGAAGCGA 2100

Qy        675 SerHisValValAsnThrGlyTrpAlaMetLeuAlaLeuIleAspAlaGlyGlnAlaLys 694
Db       2101 CCTCATGCGGTGAACACTGGTTGGGCCATGCTAGCCCTAATCGATGCAGGGCAGGCTGAG 2160

Qy        695 ArgAspSerGlnProLeuHisArgAlaAlaAlaTyrLeuIleAsnSerGlnLeuGluAsp 714
Db       2161 AGAGATCCAATTCCTTTGCATCGAGCAGCGAAGGTTTTGATCAACTTACAATCGGAAGAT 2220

Qy        715 GlyAspPheProGlnGlnGluIleMetGlyValPheAsnLysAsnCysMetIleThrTyr 734
Db       2221 GGTGAATTTCCCCAGCAAGAGATCATTGGAGTCTTCAACAAAAACTGCATGATCAGCTAC 2280

Qy        735 AlaAlaTyrArgAsnIlePheProIleTrpAlaLeuGlyGluTyrGlnSerGlnValLeu 754
Db       2281 TCCGAGTATAGAAACATCTTCCCTATTTGGGCCCTTGGGGAGTACCGTCGCCGCGTCTTG 2340

RESULT 9
US-10-425-114-24593
; Sequence 24593, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 24593
;   LENGTH: 2653
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: LIB3632-042-C4_FLI
US-10-425-114-24593

Alignment Scores:
Pred. No.:              0              Length:         2653
Score:                  3181.50        Matches:        568
Percent Similarity:     85.58%         Conservative:   79
Best Local Similarity:  75.13%         Mismatches:     97
Query Match:            76.70%         Indels:         12
DB:                     19             Gaps:           3

US-09-419-679-6 (1-757) x US-10-425-114-24593 (1-2653)

Qy          1 MetTrpLysLeuLysPheAlaGlu---GlyGlyAsnProTrpLeuArgThrLeuAsnAsn 19
Db        148 ATGTGGCGCCTGCGGATCGCGGAGGGGGGCGGGGACCCCTGGCTCCGCACGAAGAACGCC 207

Qy         20 HisValGlyArgGlnValTrpGluPheAspProLysLeuGlySerProGlnAspLeuLeu 39
Db        208 CACGTCGGCCGCCAGGTCTGGGAGTTCGAC---------GCCTCCGCCGACCCCGACCCC 258

Qy         40 GluIleGluLysAlaArgGlnAsnPheHisAspAsnArgPheThrHisLysHisSerAla 59
Db        259 GCCGTCGACGCCGCGCGCCGCGCCTTCGCCGGCAGCAGGGGCCACCTCAAGCACAGCGCC 318

Qy         60 AspLeuLeuMetArgMetGlnPheAlaArgGluAsnProThrArgGluValLeuProLys 79
Db        319 GACCTCCTCATGCGCATCCAGTTCGCCAAAGAAAATCCTCTTGAGCTTGACCTTCCAGCG 378

Qy         80 ValGlyValLysAspIleGluAspValThrGlnGluIleValThrLysThrLeuArgArg 99
Db        379 ATCAAGCTCGGTGAGCACGAAGACGTAACCGAGGAAGCCGTGTCTACTACTTTGAGAAGG 438

Qy        100 AlaValSerPheHisSerThrLeuGlnCysHisAspGlyHisTrpProGlyAspTyrGly 119
Db        439 GCAATCAGCCGTCTTTCTACTCTCCAGGCACATGATGGACACTGGCCTGGGGATTATGGT 498

Qy        120 GlyProMetPheLeuMetProGlyLeuValIleThrLeuSerIleThrGlyAlaLeuAsn 139
Db        499 GGCCCTATGTTCCTTATGCCAGGCTTGATCATAACATTGTATGTGACTGGAGCACTAAAC 558

Qy        140 ThrValLeuThrGluGluHisArgLysGluIleCysArgTyrLeuTyrAsnHisGlnAsn 159
Db        559 ACTGTCTTGTCATCAGAACACCAGAAGGAGATCCGTCGGTATCTTTATAATCATCAGAAT 618

Qy        160 LysAspGlyGlyTrpGlyLeuHisIleGluGlyProSerThrMetPheGlySerValLeu 179
Db        619 GAAGATGGTGGCTGGGGCTTGCACATTGAGGGTCCAAGCACCATGTTTGGCTCAGCACTG 678

Qy        180 SerTyrIleThrLeuArgLeuLeuGlyGluGlyProAsnAspGlyGlnGlyGluMetGlu 199
Db        679 ACCTACGTTACTTTGAGATTGCTTGGAGAGGGACCAGATAGTGGAGATGGAGCAATGGAG 738

Qy        200 LysAlaArgAspTrpIleLeuGlyHisGlyGlyAlaThrTyrIleThrSerTrpGlyLys 219
Db        739 AAAGGCCGAAACTGGATCTTAGACCATGGCGGAGCAACATATATAACATCCTGGGGGAAG 798

Qy        220 MetTrpLeuSerValLeuGlyValTyrGluTrpSerGlyAsnAsnProLeuProProGlu 239
Db        799 TTTTGGCTCTCGGTACTTGGTGTATTTGAATGGTCTGGTAACAACCCGGTACCACCAGAA 858

Qy        240 IleTrpLeuLeuProTyrMetLeuProPheHisProGlyArgMetTrpCysHisCysArg 259
Db        859 GTATGGCTACTGCCATATCTCCTGCCATTTCACCCAGGGAGGATGTGGTGTCATTGTCGA 918

Qy        260 MetValTyrLeuProMetSerTyrLeuTyrGlyLysArgPheValGlyProIleSerPro 279
Db        919 ATGGTGTATTTGCCAATGTGCTACATCTATGGGAAGAGGTTTGTTGGTCGAATCACACCA 978

Qy        280 ThrValLeuSerLeuArgLysGluLeuTyrThrValProTyrHisAspIleAspTrpAsp 299
Db        979 CTCGTGTTGGAATTAAGAAAGGAACTTTTCAAAGACCCTTACAGCAAGATTGATTGGGAC 1038

Qy        300 GlnAlaArgAsnLeuCysAlaLysGluAspLeuTyrTyrProHisProLeuValGlnAsp 319
Db       1039 AAGGCCCGCAATCTATGTGCTAAGGAAGATCTGTACTATCCACATCCATTTGTTCAAGAT 1098

Qy        320 IleLeuTrpAlaSerLeuHisLysPheLeuGluProIleLeuMetHisTrpProGlyLys 339
Db       1099 GTCTTATGGGCCACTCTCCATAAATTTGTTGAACCAGTTATGATGAGTTGGCCTGGCAGT 1158

Qy        340 ArgLeuArgGluLysAlaIleIleSerAlaLeuGluHisIleHisTyrGluAspGluAsn 359
Db       1159 AAATTGAGGGAGAAAGCTCTGGAGACAGCCATGCAACATGTTCACTATGAGGATGAGAAC 1218

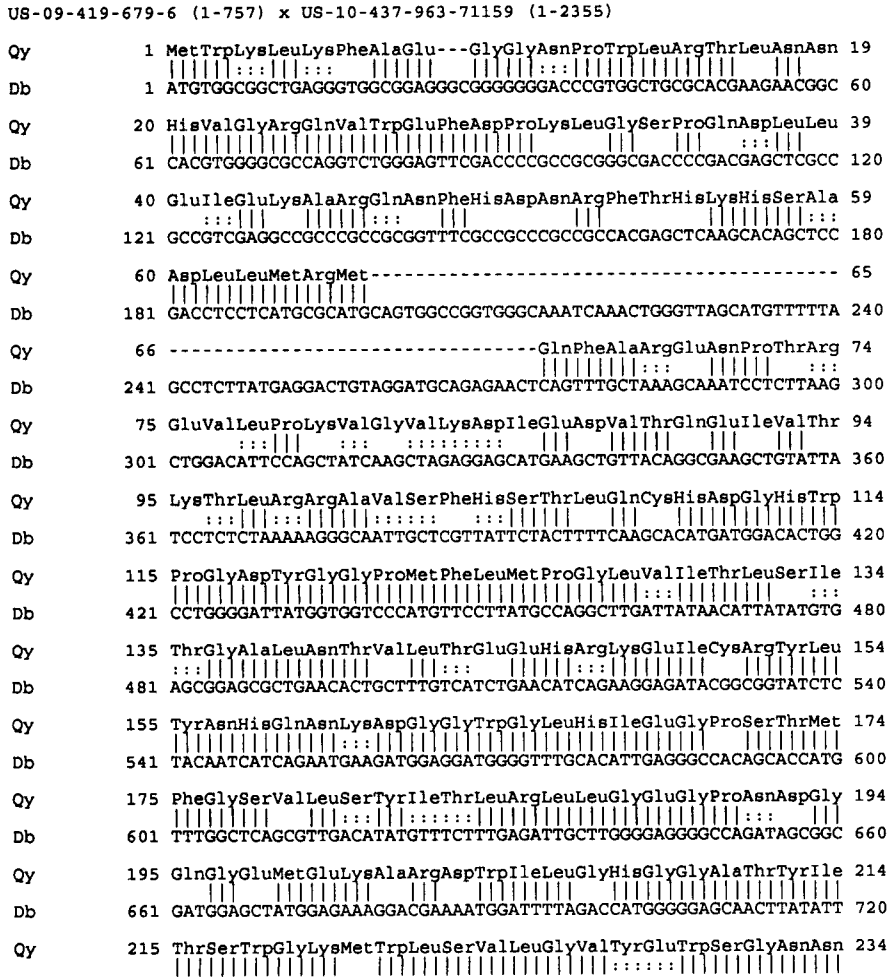
Qy        360 ThrArgTyrIleCysIleGlyProValAsnLysValLeuAsnMetLeuCysCysTrpVal 379
Db       1219 ACTCGGTACATTTGCATTGGTCCTGTAAACAAGGTATTAAATATGCTTGCTTGCTGGATT 1278

Qy        380 GluAspProAsnSerGluAlaPheLysLeuHisLeuProArgIleTyrAspTyrLeuTrp 399
```

```
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 71159
;   LENGTH: 2355
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_7165C.1
US-10-437-963-71159

Alignment Scores:
Pred. No.:                0              Length:          2355
Score:                    3234.00        Matches:         566
Percent Similarity:       84.36%         Conservative:    92
Best Local Similarity:    72.56%         Mismatches:      96
Query Match:              77.97%         Indels:          26
DB:                       20             Gaps:            2

US-09-419-679-6 (1-757) x US-10-437-963-71159 (1-2355)
```



```
Qy       100 AlaValSerPheHisSerThrLeuGlnCysHisAspGlyHisTrpProGlyAspTyrGly 119
Db       489 GCAATCAGCCGTTTCTCTACTCTCCAGGCACATGATGGACACTGGCCTGGGGATTATGGT 548

Qy       120 GlyProMetPheLeuMetProGlyLeuValIleThrLeuSerIleThrGlyAlaLeuAsn 139
Db       549 GGTCCTATGTTCCTTATGCCAGGCTTGATCATAACATTGTATGTGACTGGAGCACTAAAC 608

Qy       140 ThrValLeuThrGluGluHisArgLysGluIleCysArgTyrLeuTyrAsnHisGlnAsn 159
Db       609 ACTGTCTTGTCATTGGAACACCAGAAGGAGATCCGCCGGTATCTTTATAATCACCAGAAT 668

Qy       160 LysAspGlyGlyTrpGlyLeuHisIleGluGlyProSerThrMetPheGlySerValLeu 179
Db       669 GAAGATGGCGGCTGGGGCTTGCACATTGAGGGTCCAAGCACCATGTTCGGCTCAGCACTG 728

Qy       180 SerTyrIleThrLeuArgLeuLeuGlyGluGlyProAsnAspGlyGlnGlyGluMetGlu 199
Db       729 ACCTATGTTATTTTGAGATTGCTTGGAGAGGGACCAGATAGTGGAGATGGAGCCATGGAG 788

Qy       200 LysAlaArgAspTrpIleLeuGlyHisGlyGlyAlaThrTyrIleThrSerTrpGlyLys 219
Db       789 AAAGGTCGAAACTGGATATTAGACCATGGTGGAGCAACATATATAACATCGTGGGGGAAG 848

Qy       220 MetTrpLeuSerValLeuGlyValTyrGluTrpSerGlyAsnAsnProLeuProProGlu 239
Db       849 TTTTGGCTTTCGGTACTAGGTGTATTTGAATGGTCTGGTAACAACCCGGTGCCACCAGAA 908

Qy       240 IleTrpLeuLeuProTyrMetLeuProPheHisProGlyArgMetTrpCysHisCysArg 259
Db       909 GTATGGCTACTGCCATATCTCCTCCCATTTCACCCAGGGAGGATGTGGTGTCATTGTCGA 968

Qy       260 MetValTyrLeuProMetSerTyrLeuTyrGlyLysArgPheValGlyProIleSerPro 279
Db       969 ATGGTGTATTTGCCAATGTGCTACATTTATGGGAAGAGGTTTGTTGGCCGAATCACACCA 1028

Qy       280 ThrValLeuSerLeuArgLysGluLeuTyrThrValProTyrHisAspIleAspTrpAsp 299
Db      1029 CTTCTGTTGGAATTAAGGAAGGAACTTTTCAAAGACCCTTACAGCAAGATTGATTGGGAC 1088

Qy       300 GlnAlaArgAsnLeuCysAlaLysGluAspLeuTyrTyrProHisProLeuValGlnAsp 319
Db      1089 AAGGCCCGTAACCTATGTGCCAAGGAAGATCTGTACTACCCACACCCATTCGTTCAAGAT 1148

Qy       320 IleLeuTrpAlaSerLeuHisLysPheLeuGluProIleLeuMetHisTrpProGlyLys 339
Db      1149 GTGTTATGGGCCACTCTCCATAAATTCGTTGAACCAGTTATGATGCATTGGCCTGGCAGC 1208

Qy       340 ArgLeuArgGluLysAlaIleIleSerAlaLeuGluHisIleHisTyrGluAspGluAsn 359
Db      1209 AAATTGAGGGAGAAAGCTCTGGAAACAGTCATGCAACATGTTCATTATGAAGATGAGAAC 1268

Qy       360 ThrArgTyrIleCysIleGlyProValAsnLysValLeuAsnMetLeuCysCysTrpVal 379
Db      1269 ACTCGTTATATTTGCATTGGTCCTGTAAACAAGGTATTGAATATGCTTGCTTGCTGGATT 1328

Qy       380 GluAspProAsnSerGluAlaPheLysLeuHisLeuProArgIleTyrAspTyrLeuTrp 399
Db      1329 GAAGATCCAAACTCGGAGGCCTTCAAACTTCATATCCCACGAGTCTATGATTACTTGTGG 1388

Qy       400 IleAlaGluAspGlyMetLysMetGlnGlyTyrAsnGlySerGlnLeuTrpAspThrAla 419
Db      1389 CTTGCTGAAGATGGCATGAAGATGCAGGGTTATAATGGCAGCCAACTTTGGGATACAGCT 1448

Qy       420 PheAlaValGlnAlaIleIleAlaSerAsnLeuIleGluGluPheGlyProThrIleArg 439
Db      1449 TTCACAGTTCAAGCCATTGTGGCTACCAACCTTATTGAAGAGTTTGGTCCTACCCTTAAA 1508

Qy       440 LysAlaHisThrTyrIleLysAsnSerGlnValLeuGluAspCysProGlyAspLeuAsn 459
Db      1509 CTAGCACACAACTATATCAAGAATTCACAGGTTCTTGATGACTGCCCTGGTGATCTGAAT 1568

Qy       460 LysTrpTyrArgHisIleSerLysGlyAlaTrpProPheSerThrGlyAspHisGlyTrp 479
Db      1569 GACTGGTACCGCCACACATCTAAAGGTGCATGGCCATTCTCAACTGCTGATCATGGTTGG 1628

Qy       480 ProIleSerAspCysThrAlaGluGlyLeuLysAlaValLeuLeuLeuSerLysIleAla 499
Db      1629 CCTATATCTGATTGCACTGCTGAAGGACTAAAGGCCTCATTATTGTTATCAAGGATCTCT 1688

Qy       500 ProGluIleValGlyGluProIleAspValLysArgLeuTyrAspSerValAsnValIle 519
Db      1689 CCCAAAATTGTTGGTGAACCGATGGAAGCTAATAGATTTTATGATGCTGTCAGTTGTCTG 1748

Qy       520 LeuSerLeuGlnAsnGluAspGlyGlyPheAlaThrTyrGluLeuLysArgSerTyrAsn 539
Db      1749 ATGTCTTATATGAATGATAATGGCGGCTTCGCGACATATGAACTCACAAGATCTTATCCC 1808

Qy       540 TrpLeuGluIleIleAsnProAlaGluThrPheGlyAspIleValIleAspTyrProTyr 559
Db      1809 TGGTTGGAGCTGATCAATCCTGCTGAGACCTTTGGGGATATTGTGATTGATTACCCGTAT 1868

Qy       560 ValGluCysThrSerAlaAlaIleGlnAlaLeuAlaSerPheArgLysLeuTyrProGly 579
Db      1869 GTTGAATGTACATCAGCAGCAATTCAGGCCCTGACATCATTCAAAAAACTATACCCTGGG 1928

Qy       580 HisArgArgGluGluIleGlnHisCysIleAspLysAlaThrThrPheIleGluLysIle 599
Db      1929 CACCGCAGGAAAGAGGTGGATAACTGTATCAGCAAAGCTTCCAATTTCATCGAGAGTATT 1988

Qy       600 GlnAlaSerAspGlySerTrpTyrGlySerTrpGlyValCysPheThrTyrGlyAlaTrp 619
Db      1989 CAGAAAAGCGATGGTTCATGGTATGGCTCTTGGGCCGTTTGTTTCACATACGGCACTTGG 2048

Qy       620 PheGlyValLysGlyLeuIleAlaAlaGlyArgSerPheSerAsnCysSerSerIleArg 639
Db      2049 TTTGGTGTGAAGGGACTAATTGCTGCTGGTAGAACATTTGAGAACAGTCCTGCAATTAGA 2108

Qy       640 LysAlaCysGluPheLeuLeuSerLysGlnLeuProSerGlyGlyTrpGlyGluSerTyr 659
Db      2109 AAGGCATGCGACTTTCTGTTGTCAAAAGAACTTCCTTCCGGTGGTTGGGGAGAAAGCTAT 2168

Qy       660 LeuSerCysGlnAsnLysValTyrSerAsnLeuGluGlyAsnArgSerHisValValAsn 679
Db      2169 TTGTCATCTCAAGACCAGGTTTACACCAATCTCAAAGGCAACCGGCCTCATGCAGTGAAC 2228

Qy       680 ThrGlyTrpAlaMetLeuAlaLeuIleAspAlaGlyGlnAlaLysArgAspSerGlnPro 699
Db      2229 ACTAGTTGGGCCATGCTGGCGCTGATTGATGCAGGCCAGGCCGAGAGAGATCCAACGCCT 2288

Qy       700 LeuHisArgAlaAlaAlaTyrLeuIleAsnSerGlnLeuGluAspGlyAspProGln 719
Db      2289 CTACACCGAGCAGCAAAGGTTTTGATCAACTTGCAATCACATGATGGAGAATTTCCTCAG 2348

Qy       720 GlnGluIleMetGlyValPheAsnLysAsnCysMetIleThrTyrAlaAlaTyrArgAsn 739
Db      2349 CAAGAGATCATAGGAGTGTTCAACAAGAACTGCATGATAAGCTACTCCCAGTACAGGAAC 2408

Qy       740 IlePheProIleTrpAlaLeuGlyGluTyrGlnSerGlnValLeu 754
Db      2409 ATCTTCCCGATTTGGGCTCTGGGTGAGTACCGGTGCCGAGTCTTG 2453
```

RESULT 8
US-10-437-963-71159
; Sequence 71159, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With

```
Qy        300 GlnAlaArgAsnLeuCysAlaLysGluAspLeuTyrTyrProHisProLeuValGlnAsp 319
Db       1089 AAGGCCCGTAACCTATGTGCCAAGGAAGATCTGTACTACCCACACCCATTCGTTCAAGAT 1148

Qy        320 IleLeuTrpAlaSerLeuHisLysPheLeuGluProIleLeuMetHisTrpProGlyLys 339
Db       1149 GTGTTATGGGCCACTCTCCATAAATTCGTTGAACCAGTTATGATGCATTGGCCTGGCAGC 1208

Qy        340 ArgLeuArgGluLysAlaIleIleSerAlaLeuGluHisIleHisTyrGluAspGluAsn 359
Db       1209 AAATTGAGGGAGAAAGCTCTGGAAACAGTCATGCAACATGTTCATTATGAAGATGAGAAC 1268

Qy        360 ThrArgTyrIleCysIleGlyProValAsnLysValLeuAsnMetLeuCysCysTrpVal 379
Db       1269 ACTCGTTATATTTGCATTGGTCCTGTAAACAAGGTATTGAATATGCTTGCTTGCTGGATT 1328

Qy        380 GluAspProAsnSerGluAlaPheLysLeuHisLeuProArgIleTyrAspTyrLeuTrp 399
Db       1329 GAAGATCCAAACTCGGAGGCCTTCAAACTTCATATCCCACGAGTCTATGATTACTTGTGG 1388

Qy        400 IleAlaGluAspGlyMetLysMetGlnGlyTyrAsnGlySerGlnLeuTrpAspThrAla 419
Db       1389 CTTGCTGAAGATGGCATGAAGATGCAGGGTTATAATGGCAGCCAACTTTGGGATACAGCT 1448

Qy        420 PheAlaValGlnAlaIleIleAlaSerAsnLeuIleGluGluPheGlyProThrIleArg 439
Db       1449 TTCACAGTTCAAGCCATTGTGGCTACCAACCTTATTGAAGAGTTTGGTCCTACCCTTAAA 1508

Qy        440 LysAlaHisThrTyrIleLysAsnSerGlnValLeuGluAspCysProGlyAspLeuAsn 459
Db       1509 CTAGCACACAACTATATCAAGAATTCACAGGTTCTTGATGACTGCCCTGGTGATCTGAAT 1568

Qy        460 LysTrpTyrArgHisIleSerLysGlyAlaTrpProPheSerThrGlyAspHisGlyTrp 479
Db       1569 GACTGGTACCGCCACACATCTAAAGGTGCATGGCCATTCTCAACTGCTGATCATGGTTGG 1628

Qy        480 ProIleSerAspCysThrAlaGluGlyLeuLysAlaValLeuLeuLeuSerLysIleAla 499
Db       1629 CCTATATCTGATTGCACTGCTGAAGGACTAAAGGCCTCATTATTGTTATCAAGGATCTCT 1688

Qy        500 ProGluIleValGlyGluProIleAspValLysArgLeuTyrAspSerValAsnValIle 519
Db       1689 CCCAAAATTGTTGGTGAACCGATGGAAGCTAATAGATTTTATGATGCTGTCAGTTGTCTG 1748

Qy        520 LeuSerLeuGlnAsnGluAspGlyGlyPheAlaThrTyrGluLeuLysArgSerTyrAsn 539
Db       1749 ATGTCTTATATGAATGATAATGGCGGCTTCGCGACATATGAACTCACAAGATCTTATCCC 1808

Qy        540 TrpLeuGluIleIleAsnProAlaGluThrPheGlyAspIleValIleAspTyrProTyr 559
Db       1809 TGGTTGGAGCTGATCAATCCTGCTGAGACCTTTGGGGATATTGTGATTGATTACCCGTAT 1868

Qy        560 ValGluCysThrSerAlaAlaIleGlnAlaLeuAlaSerPheArgLysLeuTyrProGly 579
Db       1869 GTTGAATGTACATCAGCAGCAATTCAGGCCCTGACATCATTCAAAAAACTATACCCTGGG 1928

Qy        580 HisArgArgGluGluIleGlnHisCysIleAspLysAlaThrThrPheIleGluLysIle 599
Db       1929 CACCGCAGGAAAGAGGTGGATAACTGTATCAGCAAAGCTTCCAATTTCATCGAGAGTATT 1988

Qy        600 GlnAlaSerAspGlySerTrpTyrGlySerTrpGlyValCysPheThrTyrGlyAlaTrp 619
Db       1989 CAGAAAAGCGATGGTTCATGGTATGGCTCTTGGGCCGTTTGTTTCACATACGGCACTTGG 2048

Qy        620 PheGlyValLysGlyLeuIleAlaAlaGlyArgSerPheSerAsnCysSerSerIleArg 639
Db       2049 TTTGGTGTGAAGGGACTAATTGCTGCTGGTAGAACATTTGAGAACAGTCCTGCAATTAGA 2108

Qy        640 LysAlaCysGluPheLeuLeuSerLysGlnLeuProSerGlyGlyTrpGlyGluSerTyr 659
Db       2109 AAGGCATGCGACTTTCTGTTGTCAAAAGAACTTCCTTCCGGTGGTTGGGGAGAAAGCTAT 2168

Qy        660 LeuSerCysGlnAsnLysValTyrSerAsnLeuGluGlyAsnArgSerHisValValAsn 679
Db       2169 TTGTCATCTCAAGACCAGGTTTACACCAATCTCAAAGGCAACCGGCCTCATGCAGTGAAC 2228

Qy        680 ThrGlyTrpAlaMetLeuAlaLeuIleAspAlaGlyGlnAlaLysArgAspSerGlnPro 699
Db       2229 ACTAGTTGGGCCATGCTGGCGCTGATTGATGCAGGCCAGGCCGAGAGAGATCCAACGCCT 2288

Qy        700 LeuHisArgAlaAlaAlaTyrLeuIleAsnSerGlnLeuGluAspGlyAspPheProGln 719
Db       2289 CTACACCGAGCAGCAAAGGTTTTGATCAACTTGCAATCACATGATGGAGAATTTCCTCAG 2348

Qy        720 GlnGluIleMetGlyValPheAsnLysAsnCysMetIleThrTyrAlaAlaTyrArgAsn 739
Db       2349 CAAGAGATCATAGGAGTGTTCAACAAGAACTGCATGATAAGCTACTCCCAGTACAGGAAC 2408

Qy        740 IlePheProIleTrpAlaLeuGlyGluTyrGlnSerGlnValLeu 754
Db       2409 ATCTTCCCGATTTGGGCTCTGGGTGAGTACCGGTGCCGAGTCTTG 2453
```

RESULT 7
US-10-425-115-90182
; Sequence 90182, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 90182
; LENGTH: 2786
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_182249C.1
US-10-425-115-90182

Alignment Scores:

Pred. No.:	0	Length:	2786
Score:	3237.00	Matches:	571
Percent Similarity:	86.75%	Conservative:	84
Best Local Similarity:	75.63%	Mismatches:	96
Query Match:	78.04%	Indels:	4
DB:	21	Gaps:	3

US-09-419-679-6 (1-757) x US-10-425-115-90182 (1-2786)

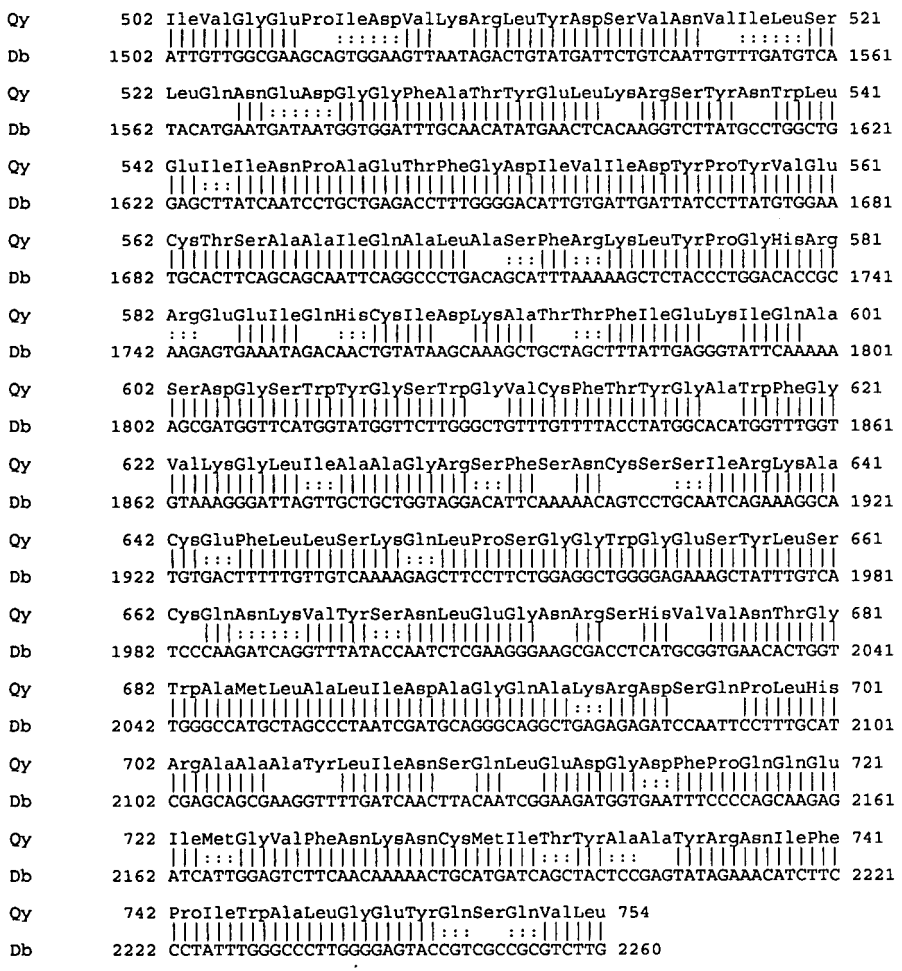
```
Qy          1 MetTrpLysLeuLysPheAlaGlu---GlyGlyAsnProTrpLeuArgThrLeuAsnAsn 19
Db        198 ATGTGGCGCCTGCGGATCGCGGAGGGGGGCGGGGACCCCTGGCTACGCACGAAGAACGCC 257

Qy         20 HisValGlyArgGlnValTrpGluPheAspProLysLeuGlySerProGlnAspLeuLeu 39
Db        258 CACGTCGGCCGCCAGGTGTGGGAGTTCGAC---GCCGCCGCCGACCCCGACCCC------ 308

Qy         40 GluIleGluLysAlaArgGlnAsnPheHisAspAsnArgPheThrHisLysHisSerAla 59
Db        309 GCCGTCGACGCCGCGCGCCGCGCCTTCGTCGCCACCAGGGGCCACCTCAAGCACAGCGCC 368

Qy         60 AspLeuLeuMetArgMetGlnPheAlaArgGluAsnProThrArgGluValLeuProLys 79
Db        369 GACCTCCTCATGCGCATCCAGTTCGCCAAAGAAAACTCGATTGAGCTTCACCTTCCAGGC 428

Qy         80 ValGlyValLysAspIleGluAspValThrGlnGluIleValThrLysThrLeuArgArg 99
Db        429 ATCAAGCTCGGTGAGCATGAAGATGTGACCGAGGAAGCTGTGTTGACTACATTGAAAAGG 488
```

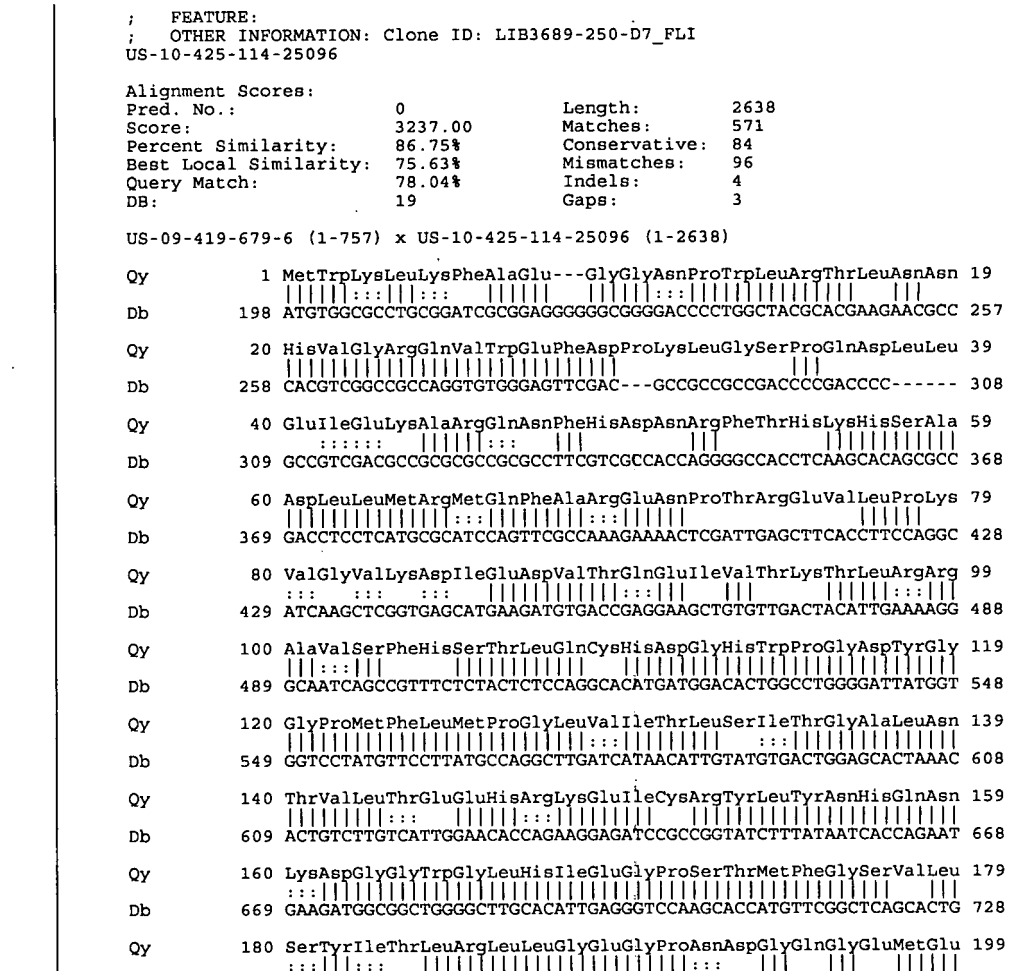


RESULT 6
US-10-425-114-25096
; Sequence 25096, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128.
; SEQ ID NO 25096
; LENGTH: 2638
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3689-250-D7_FLI
US-10-425-114-25096

Alignment Scores:

Pred. No.:	0	Length:	2638
Score:	3237.00	Matches:	571
Percent Similarity:	86.75%	Conservative:	84
Best Local Similarity:	75.63%	Mismatches:	96
Query Match:	78.04%	Indels:	4
DB:	19	Gaps:	3

US-09-419-679-6 (1-757) x US-10-425-114-25096 (1-2638)



```
Qy        720 GlnGluIleMetGlyValPheAsnLysAsnCysMetIleThrTyrAlaAlaTyrArgAsn 739
Db       2353 CAAGAGATCATAGGAGTGTTCAACAAGAACTGCATGATCAGTTACTCCCAGTACAGGAAC 2412

Qy        740 IlePheProIleTrpAlaLeuGlyGluTyrGlnSerGlnValLeu 754
Db       2413 ACCTTCCCGATCTGGGCTCTGGGAGAGTACCGGTGCCGAGTCCTG 2457
```

RESULT 5
US-10-437-963-71137
; Sequence 71137, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 71137
; LENGTH: 4120
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_7163C.1
US-10-437-963-71137

Alignment Scores:

Pred. No.:	0	Length:	4120
Score:	3240.50	Matches:	564
Percent Similarity:	87.12%	Conservative:	92
Best Local Similarity:	74.90%	Mismatches:	96
Query Match:	78.12%	Indels:	1
DB:	20	Gaps:	1

US-09-419-679-6 (1-757) x US-10-437-963-71137 (1-4120)

```
Qy          3 LysLeuLysPheAlaGlu---GlyGlyAsnProTrpLeuArgThrLeuAsnAsnHisVal 21
Db          2 CGGCTGAGGGTGGCGGAGGGCGGGGGGGACCCGTGGCTGCGCACGAAAAACGGCCACGTG 61

Qy         22 GlyArgGlnValTrpGluPheAspProLysLeuGlySerProGlnAspLeuLeuGluIle 41
Db         62 GGGCGCCAGGTCTGGGAGTTCGACCCCGCCGCGGGCGACCCCGACGAGCTCGCCGCCGTC 121

Qy         42 GluLysAlaArgGlnAsnPheHisAspAsnArgPheThrHisLysHisSerAlaAspLeu 61
Db        122 GAGGCCGCCCGCCGCGGTTTCGCCGCCCGCCGCCACGAGCTCAAGCACAGCTCCGACCTC 181

Qy         62 LeuMetArgMetGlnPheAlaArgGluAsnProThrArgGluValLeuProLysValGly 81
Db        182 CTCATGCGCATGCAGTTTGCTAAAGCAAATCCTCTTAAGCTGGACATTCCAGCTATCAAG 241

Qy         82 ValLysAspIleGluAspValThrGlnGluIleValThrLysThrLeuArgArgAlaVal 101
Db        242 CTAGAGGAGCATGAAGCTGTTACAGGCGAAGCTGTATTATCCTCTCTAAAAAGGGCAATT 301

Qy        102 SerPheHisSerThrLeuGlnCysHisAspGlyHisTrpProGlyAspTyrGlyGlyPro 121
Db        302 GCTCGTTATTCTACTTTTCAAGCACATGATGGACACTGGCCTGGGGATTATGGTGGTCCC 361

Qy        122 MetPheLeuMetProGlyLeuValIleThrLeuSerIleThrGlyAlaLeuAsnThrVal 141
Db        362 ATGTTCCTTATGCCAGGCTTGATTATAACATTATATGTGAGCGGAGCGCTGAACACTGCT 421

Qy        142 LeuThrGluGluHisArgLysGluIleCysArgTyrLeuTyrAsnHisGlnAsnLysAsp 161
Db        422 TTGTCATCTGAACATCAGAAGGAGATACGGCGGTATCTCTACAATCATCAGAATGAAGAT 481

Qy        162 GlyGlyTrpGlyLeuHisIleGluGlyProSerThrMetPheGlySerValLeuSerTyr 181
Db        482 GGAGGATGGGGTTTGCACATTGAGGGCCACAGCACCATGTTTGGCTCAGCGTTGACATAT 541

Qy        182 IleThrLeuArgLeuLeuGlyGluGlyProAsnAspGlyGlnGlyGluMetGluLysAla 201
Db        542 GTTTCTTTGAGATTGCTTGGGGAGGGGCCAGATAGCGGCGATGGAGCTATGGAGAAAGGA 601

Qy        202 ArgAspTrpIleLeuGlyHisGlyGlyAlaThrTyrIleThrSerTrpGlyLysMetTrp 221
Db        602 CGAAAATGGATTTTAGACCATGGGGGAGCAACTTATATTACATCATGGGGCAAGTTTTGG 661

Qy        222 LeuSerValLeuGlyValTyrGluTrpSerGlyAsnAsnProLeuProProGluIleTrp 241
Db        662 CTTTCGGTGCTTGGTGTATTTGACTGGTCTGGCAACAACCCAGTGCCACCAGAAATATGG 721

Qy        242 LeuLeuProTyrMetLeuProPheHisProGlyArgMetTrpCysHisCysArgMetVal 261
Db        722 TTGTTGCCATATTTCCTGCCGATTCATCCAGGGCGAATGTGGTGTCATTGCCGGATGGTT 781

Qy        262 TyrLeuProMetSerTyrLeuTyrGlyLysArgPheValGlyProIleSerProThrVal 281
Db        782 TATTTGCCTATGTGTTACATTTATGGAAAGAGGTTTGTGGGCCCAGTTACACCAATTATA 841

Qy        282 LeuSerLeuArgLysGluLeuTyrThrValProTyrHisAspIleAspTrpAspGlnAla 301
Db        842 TTGGAATTAAGAAAGGAACTCTACGAAGTACCCTACAATGAAGTTGATTGGGACAAGGCT 901

Qy        302 ArgAsnLeuCysAlaLysGluAspLeuTyrTyrProHisProLeuValGlnAspIleLeu 321
Db        902 CGCAATCTATGTGCTAAGGAAGATCTGTACTATCCACATCCATTCGTGCAGGATGTATTA 961

Qy        322 TrpAlaSerLeuHisLysPheLeuGluProIleLeuMetHisTrpProGlyLysArgLeu 341
Db        962 TGGGCCACTCTCCACAAATTTGTTGAACCAGCTATGTTGCGTTGGCCTGGGAACAAATTG 1021

Qy        342 ArgGluLysAlaIleIleSerAlaLeuGluHisIleHisTyrGluAspGluAsnThrArg 361
Db       1022 AGGGAGAAAGCTTTGGACACTGTCATGCAGCATATTCATTATGAAGATGAGAACACCCGA 1081

Qy        362 TyrIleCysIleGlyProValAsnLysValLeuAsnMetLeuCysCysTrpValGluAsp 381
Db       1082 TATATTTGCATTGGTCCAGTAAACAAGGTATTAAATATGCTTGCTTGCTGGATTGAAGAT 1141

Qy        382 ProAsnSerGluAlaPheLysLeuHisLeuProArgIleTyrAspTyrLeuTrpIleAla 401
Db       1142 CCAAACTCAGAGGCATTCAAACTCCACATTGCAAGAGTCCACGATTACCTATGGATTGCA 1201

Qy        402 GluAspGlyMetLysMetGlnGlyTyrAsnGlySerGlnLeuTrpAspThrAlaPheAla 421
Db       1202 GAAGATGGCATGAAAATGCAGGGTTATAATGGAAGCCAGCTGTGGGACACAGCTTTCACA 1261

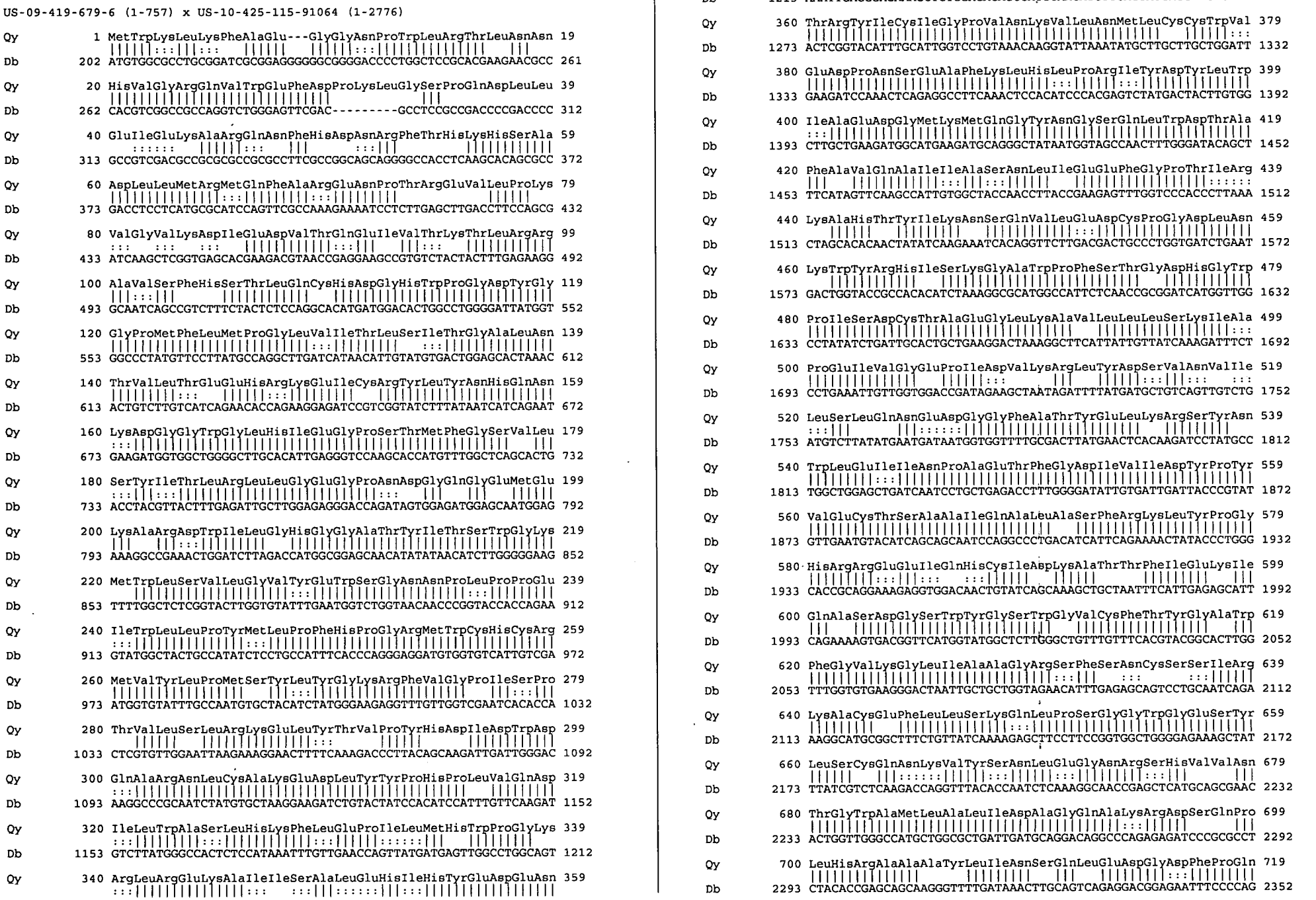
Qy        422 ValGlnAlaIleIleAlaSerAsnLeuIleGluGluPheGlyProThrIleArgLysAla 441
Db       1262 GTTCAAGCTATAGTGGCTACTGGCCTCATTGAAGAATTTGGTCCTACTCTTAAACTAGCA 1321

Qy        442 HisThrTyrIleLysAsnSerGlnValLeuGluAspCysProGlyAspLeuAsnLysTrp 461
Db       1322 CATGGCTACATAAAGAAAACGCAGGTTATCGATGACTGCCCTGGAGATCTTAGTCAGTGG 1381

Qy        462 TyrArgHisIleSerLysGlyAlaTrpProPheSerThrGlyAspHisGlyTrpProIle 481
Db       1382 TACCGCCACATATCTAAAGGTGCATGGCCCTTTTCTACTGCTGATCATGGTTGGCCTATA 1441

Qy        482 SerAspCysThrAlaGluGlyLeuLysAlaValLeuLeuLeuSerLysIleAlaProGlu 501
Db       1442 TCAGATTGCACTGCAGAAGGACTTAAGGCGGCATTATTGCTATCGAAGATTTCTCCAGAT 1501
```

US-09-419-679-6 (1-757) x US-10-425-115-91064 (1-2776)



```
Qy       181 TyrIleThrLeuArgLeuLeuGlyGluGlyProAsnAspGlyGlnGlyGluMetGluLys 200
Db       541 TATGTTACTCTAAGGTTGCTTGGAGAAGGACCTAACGATGGAGATGGAGATATGGAGAAA 600

Qy       201 AlaArgAspTrpIleLeuGlyHisGlyGlyAlaThrTyrIleThrSerTrpGlyLysMet 220
Db       601 GGACGAGACTGGATACTAAATCATGGTGGTGCTACCAATATTACATCTTGGGGGAAAATG 660

Qy       221 TrpLeuSerValLeuGlyValTyrGluTrpSerGlyAsnAsnProLeuProProGluIle 240
Db       661 TGGCTATCGGTACTTGGAGCTTTTGAATGGTCCGGAAATAACCCACTGCCACCTGAGATA 720

Qy       241 TrpLeuLeuProTyrMetLeuProPheHisProGlyArgMetTrpCysHisCysArgMet 260
Db       721 TGGCTTCTCCCATATTTCCTGCCAATACATCCAGGAAGGATGTGGTGCCATTGTCGAATG 780

Qy       261 ValTyrLeuProMetSerTyrLeuTyrGlyLysArgPheValGlyProIleSerProThr 280
Db       781 GTGTACTTGCCGATGTCGTATTTGTATGGAAAAAGGTTTGTGGGTCCCATAACGTCCACT 840

Qy       281 ValLeuSerLeuArgLysGluLeuTyrThrValProTyrHisAspIleAspTrpAspGln 300
Db       841 GTTTTATCACTGAGAAAGGAGCTTTTCACAGTACCATATCATGAAGTCAACTGGAATGAA 900

Qy       301 AlaArgAsnLeuCysAlaLysGluAspLeuTyrTyrProHisProLeuValGlnAspIle 320
Db       901 GCACGCAACCTTTGCGCAAAGGAGGATTTATACTACCCACATCCACTTGTGCAAGATATT 960

Qy       321 LeuTrpAlaSerLeuHisLysPheLeuGluProIleLeuMetHisTrpProGlyLysArg 340
Db       961 CTTTGGGCATCACTTCATAAGATTGTTGAGCCTGTTCTGATGCGATGGCCTGGTGCAAAT 1020

Qy       341 LeuArgGluLysAlaIleIleSerAlaLeuGluHisIleHisTyrGluAspGluAsnThr 360
Db      1021 TTGAGAGAAAAGGCTATAAGAACCGCAATAGAACATATTCATTATGAAGATGAGAATACT 1080

Qy       361 ArgTyrIleCysIleGlyProValAsnLysValLeuAsnMetLeuCysCysTrpValGlu 380
Db      1081 AGGTACATCTGCATAGGTCCCGTGAACAAGGTATTAAATATGCTTTGCTGTTGGGTAGAA 1140

Qy       381 AspProAsnSerGluAlaPheLysLeuHisLeuProArgIleTyrAspTyrLeuTrpIle 400
Db      1141 GACCCAAACTCAGAGGCTTTCAAGTTGCACCTACCAAGAATCCATGACTTTCTCTGGTTA 1200

Qy       401 AlaGluAspGlyMetLysMetGlnGlyTyrAsnGlySerGlnLeuTrpAspThrAlaPhe 420
Db      1201 GCTGAAGATGGAATGAAGATGCAGGGTTATAACGGAAGCCAGCTATGGGATACAGGTTTT 1260

Qy       421 AlaValGlnAlaIleIleAlaSerAsnLeuIleGluGluPheGlyProThrIleArgLys 440
Db      1261 GCTATTCAAGCGATTTTGGCAACTAACCTCGTCGAAGAATATGGGCCCGTTTTGGAAAAA 1320

Qy       441 AlaHisThrTyrIleLysAsnSerGlnValLeuGluAspCysProGlyAspLeuAsnLys 460
Db      1321 GCACATTCATTTGTCAAGAATTCCCAGGTGTTAGAAGACTGCCCTGGAGATCTGAATTAC 1380

Qy       461 TrpTyrArgHisIleSerLysGlyAlaTrpProPheSerThrGlyAspHisGlyTrpPro 480
Db      1381 TGGTATCGCCACATTTCTAAAGGGGCTTGGCCTTTCTCAACTGCAGATCACGGTTGGCCC 1440

Qy       481 IleSerAspCysThrAlaGluGlyLeuLysAlaValLeuLeuLeuSerLysIleAlaPro 500
Db      1441 ATCTCTGACTGCACCGCAGAAGGACTGAAAGCTGCTCTTTTGCTATCCAAAGTTCCCAAG 1500

Qy       501 GluIleValGlyGluProIleAspValLysArgLeuTyrAspSerValAsnValIleLeu 520
Db      1501 GCGATTGTTGGTGAACCAATAGATGCAAAACGGTTATATGAAGCTGTTAATGTTATCATT 1560

Qy       521 SerLeuGlnAsnGluAspGlyGlyPheAlaThrTyrGluLeuLysArgSerTyrAsnTrp 540
Db      1561 TCTTTACAGAATGCAGATGGAGGCCTCGCAACATATGAGCTCACCAGGTCATACCCTTGG 1620

Qy       541 LeuGluIleIleAsnProAlaGluThrPheGlyAspIleValIleAspTyrProTyrVal 560
Db      1621 TTAGAGCTAATCAACCCAGCAGAAACCTTTGGCGATATTGTTATTGATTATCCTTACGTG 1680

Qy       561 GluCysThrSerAlaAlaIleGlnAlaLeuAlaSerPheArgLysLeuTyrProGlyHis 580
Db      1681 GAATGTACATCAGCTGCTATCCAAGCTTTGATATCATTTCGAAAGCTGTATCCTGGTCAT 1740

Qy       581 ArgArgGluGluIleGlnHisCysIleAspLysAlaThrThrPheIleGluLysIleGln 600
Db      1741 CGAAAGAAGGAAGTAGATGAGTGCATTGAGAAGGCGGTTAAGTTCATTGAATCCATTCAA 1800

Qy       601 AlaSerAspGlySerTrpTyrGlySerTrpGlyValCysPheThrTyrGlyAlaTrpPhe 620
Db      1801 GCAGCAGATGGCTCATGGTATGGATCATGGGCTGTTTGCTTCACGTATGGTACGTGGTTT 1860

Qy       621 GlyValLysGlyLeuIleAlaAlaGlyArgSerPheSerAsnCysSerSerIleArgLys 640
Db      1861 GGAGTGAAAGGGCTGGTAGCTGTTGGAAAAACATTGAAAAACTCTCCACATGTTGCTAAA 1920

Qy       641 AlaCysGluPheLeuLeuSerLysGlnLeuProSerGlyGlyTrpGlyGluSerTyrLeu 660
Db      1921 GCTTGTGAATTTCTATTGTCGAAACAACAACCTTCGGGCGGCTGGGGAGAAAGCTATCTT 1980

Qy       661 SerCysGlnAsnLysValTyrSerAsnLeuGluGlyAsnArgSerHisValValAsnThr 680
Db      1981 TCATGTCAAGACAAGGTCTATTCAAACCTTGATGGCAACAGATCTCACGTCGTGAATACA 2040

Qy       681 GlyTrpAlaMetLeuAlaLeuIleAspAlaGlyGlnAlaLysArgAspSerGlnProLeu 700
Db      2041 GCATGGGCTATGCTCGCACTCATTGGTGCTGGGCAAGCTGAGGTAGACCGGAAACCACTA 2100

Qy       701 HisArgAlaAlaAlaTyrLeuIleAsnSerGlnLeuGluAspGlyAspPheProGlnGln 720
Db      2101 CACCGGGCTGCAAGATACTTGATTAATGCTCAAATGGAGAATGGTGATTTTCCACAACAG 2160

Qy       721 GluIleMetGlyValPheAsnLysAsnCysMetIleThrTyrAlaAlaTyrArgAsnIle 740
Db      2161 GAAATAATGGGAGTCTTCAATAGGAACTGCATGATAACATATGCCGCGTATCGAAACATT 2220

Qy       741 PheProIleTrpAlaLeuGlyGluTyrGlnSerGlnValLeu 754
Db      2221 TTTCCGATATGGGCTTTGGGGGAGTACCGTTGTCAGGTATTA 2262
```

RESULT 4
US-10-425-115-91064
; Sequence 91064, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 91064
; LENGTH: 2776
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_183051C.1
US-10-425-115-91064

Alignment Scores:

Pred. No.:	0	Length:	2776
Score:	3248.00	Matches:	575
Percent Similarity:	86.75%	Conservative:	80
Best Local Similarity:	76.16%	Mismatches:	96
Query Match:	78.30%	Indels:	4
DB:	21	Gaps:	2

```
Qy       401 AlaGluAspGlyMetLysMetGlnGlyTyrAsnGlySerGlnLeuTrpAspThrAlaPhe 420
Db      1201 GCTGAAGATGGAATGAAGATGCAGGGTTATAACGGAAGCCAGCTATGGGATACAGGTTTT 1260

Qy       421 AlaValGlnAlaIleIleAlaSerAsnLeuIleGluGluPheGlyProThrIleArgLys 440
Db      1261 GCTATTCAAGCGATTTTGGCAACTAACCTCGTCGAAGAATATGGGCCCGTTTTGGAAAAA 1320

Qy       441 AlaHisThrTyrIleLysAsnSerGlnValLeuGluAspCysProGlyAspLeuAsnLys 460
Db      1321 GCACATTCATTTGTCAAGAATTCCCAGGTGTTAGAAGACTGCCCTGGAGATCTGAATTAC 1380

Qy       461 TrpTyrArgHisIleSerLysGlyAlaTrpProPheSerThrGlyAspHisGlyTrpPro 480
Db      1381 TGGTATCGCCACATTTCTAAAGGGGCTTGGCCTTTCTCAACTGCAGATCACGGTTGGCCC 1440

Qy       481 IleSerAspCysThrAlaGluGlyLeuLysAlaValLeuLeuLeuSerLysIleAlaPro 500
Db      1441 ATCTCTGACTGCACCGCAGAAGGACTGAAAGCTGCTCTTTTGCTATCCAAAGTTCCCAAG 1500

Qy       501 GluIleValGlyGluProIleAspValLysArgLeuTyrAspSerValAsnValIleLeu 520
Db      1501 GCGATTGTTGGTGAACCAATAGATGCAAAACGGTTATATGAAGCTGTTAATGTTATCATT 1560

Qy       521 SerLeuGlnAsnGluAspGlyGlyPheAlaThrTyrGluLeuLysArgSerTyrAsnTrp 540
Db      1561 TCTTTACAGAATGCAGATGGAGGCCTCGCAACATATGAGCTCACCAGGTCATACCCTTGG 1620

Qy       541 LeuGluIleIleAsnProAlaGluThrPheGlyAspIleValIleAspTyrProTyrVal 560
Db      1621 TTAGAGCTAATCAACCCAGCAGAAACCTTTGGCGATATTGTTATTGATTATCCTTACGTG 1680

Qy       561 GluCysThrSerAlaAlaIleGlnAlaLeuAlaSerPheArgLysLeuTyrProGlyHis 580
Db      1681 GAATGTACATCAGCTGCTATCCAAGCTTTGATATCATTTCGAAAGCTGTATCCTGGTCAT 1740

Qy       581 ArgArgGluGluIleGlnHisCysIleAspLysAlaThrThrPheIleGluLysIleGln 600
Db      1741 CGAAAGAAGGAAGTAGATGAGTGCATTGAGAAGGCGGTTAAGTTCATTGAATCCATTCAA 1800

Qy       601 AlaSerAspGlySerTrpTyrGlySerTrpGlyValCysPheThrTyrGlyAlaTrpPhe 620
Db      1801 GCAGCAGATGGCTCATGGTATGGATCATGGGCTGTTTGCTTCACGTATGGTACGTGGTTT 1860

Qy       621 GlyValLysGlyLeuIleAlaAlaGlyArgSerPheSerAsnCysSerSerIleArgLys 640
Db      1861 GGAGTGAAAGGGCTGGTAGCTGTTGGAAAAACATTGAAAAACTCTCCACATGTTGCTAAA 1920

Qy       641 AlaCysGluPheLeuLeuSerLysGlnLeuProSerGlyGlyTrpGlyGluSerTyrLeu 660
Db      1921 GCTTGTGAATTTCTATTGTCGAAACAACAACCTTCGGGCGGCTGGGGAGAAAGCTATCTT 1980

Qy       661 SerCysGlnAsnLysValTyrSerAsnLeuGluGlyAsnArgSerHisValValAsnThr 680
Db      1981 TCATGTCAAGACAAGGTCTATTCAAACCTTGATGGCAACAGATCTCACGTCGTGAATACA 2040

Qy       681 GlyTrpAlaMetLeuAlaLeuIleAspAlaGlyGlnAlaLysArgAspSerGlnProLeu 700
Db      2041 GCATGGGCTATGCTCGCACTCATTGGTGCTGGGCAAGCTGAGGTAGACCGGAAACCACTA 2100

Qy       701 HisArgAlaAlaAlaTyrLeuIleAsnSerGlnLeuGluAspGlyAspPheProGlnGln 720
Db      2101 CACCGGGCTGCAAGATACTTGATTAATGCTCAAATGGAGAATGGTGATTTTCCACAACAG 2160

Qy       721 GluIleMetGlyValPheAsnLysAsnCysMetIleThrTyrAlaAlaTyrArgAsnIle 740
Db      2161 GAAATAATGGGAGTCTTCAATAGGAACTGCATGATAACATATGCCGCGTATCGAAACATT 2220

Qy       741 PheProIleTrpAlaLeuGlyGluTyrGlnSerGlnValLeu 754
Db      2221 TTTCCGATATGGGCTTTGGGGGAGTACCGTTGTCAGGTATTA 2262
```

```
RESULT 3
US-09-938-842A-833
; Sequence 833, Application US/09938842A
; Publication No. US20040009476A9
; GENERAL INFORMATION:
;  APPLICANT: Harper, Jeff
;  APPLICANT:  Kreps, Joel
;  APPLICANT:  Wang, Xun
;  APPLICANT:  Zhu, Tong
;  TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
;  TITLE OF INVENTION:  SAME, AND METHODS OF USE
;  FILE REFERENCE: SCRIP1300-3
;  CURRENT APPLICATION NUMBER: US/09/938,842A
;  CURRENT FILING DATE:  2001-08-24
;  PRIOR APPLICATION NUMBER: US 60/227,866
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: US 60/264,647
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/300,111
;  PRIOR FILING DATE: 2001-06-22
;  NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 833
;   LENGTH: 2280
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-938-842A-833

Alignment Scores:
Pred. No.:                0                Length:          2280
Score:                    3480.00          Matches:         607
Percent Similarity:       90.72%           Conservative:    77
Best Local Similarity:    80.50%           Mismatches:      70
Query Match:              83.90%           Indels:          0
DB:                       11               Gaps:            0

US-09-419-679-6 (1-757) x US-09-938-842A-833 (1-2280)

Qy         1 MetTrpLysLeuLysPheAlaGluGlyGlyAsnProTrpLeuArgThrLeuAsnAsnHis 20
Db         1 ATGTGGAAACTGAAGATCGCGGAAGGAGGTAGTCCATGGCTTAGAACCACCAATAATCAC 60

Qy        21 ValGlyArgGlnValTrpGluPheAspProLysLeuGlySerProGlnAspLeuLeuGlu 40
Db        61 GTCGGAAGACAGTTTTGGGAGTTCGATCCGAATCTCGGTACTCCTGAGGATCTCGCCGCC 120

Qy        41 IleGluLysAlaArgGlnAsnPheHisAspAsnArgPheThrHisLysHisSerAlaAsp 60
Db       121 GTCGAAGAAGCTAGGAAGTCTTTTTCGGATAATCGATTCGTGCAGAAACATAGCGCCGAT 180

Qy        61 LeuLeuMetArgMetGlnPheAlaArgGluAsnProThrArgGluValLeuProLysVal 80
Db       181 CTGCTTATGCGCCTTCAGTTTTCAAGAGAAAATTTGATTAGCCCAGTTTTACCTCAAGTC 240

Qy        81 GlyValLysAspIleGluAspValThrGlnGluIleValThrLysThrLeuArgArgAla 100
Db       241 AAAATCGAAGACACTGATGATGTTACAGAGGAGATGGTGGAAACCACGTTAAAGAGGGGT 300

Qy       101 ValSerPheHisSerThrLeuGlnCysHisAspGlyHisTrpProGlyAspTyrGlyGly 120
Db       301 CTAGATTTCTATTCAACTATACAGGCACACGACGGGCACTGGCCAGGTGATTATGGTGGT 360

Qy       121 ProMetPheLeuMetProGlyLeuValIleThrLeuSerIleThrGlyAlaLeuAsnThr 140
Db       361 CCTATGTTTCTTCTCCCAGGACTGATAATTACACTCTCCATAACTGGAGCACTGAATACA 420

Qy       141 ValLeuThrGluGluHisArgLysGluIleCysArgTyrLeuTyrAsnHisGlnAsnLys 160
Db       421 GTATTGTCGGAACAACATAAACAAGAAATGCGCCGTTATCTCTATAATCACCAGAATGAG 480

Qy       161 AspGlyGlyTrpGlyLeuHisIleGluGlyProSerThrMetPheGlySerValLeuSer 180
Db       481 GACGGAGGTTGGGGTTTACATATTGAGGGCCCTAGCACCATGTTTGGGTCTGTGTTGAAC 540
```

```
Db       1825 GCTTCAGATGGATCATGGTATGGTTCTTGGGGAGTTTGCTTCACTTACGGTGCTTGGTTT 1884

Qy        621 GlyValLysGlyLeuIleAlaAlaGlyArgSerPheSerAsnCysSerSerIleArgLys 640

Db       1885 GGGGTAAAAGGTCTGATTGCTGCTGGAAGGAGTTTCAGTAATTGCTCAAGCATCCGTAAA 1944

Qy        641 AlaCysGluPheLeuLeuSerLysGlnLeuProSerGlyGlyTrpGlyGluSerTyrLeu 660

Db       1945 GCTTGTGAATTTCTGCTGTCCAAGCAGCTTCCTTCTGGTGGCTGGGGAGAGAGTTATCTG 2004

Qy        661 SerCysGlnAsnLysValTyrSerAsnLeuGluGlyAsnArgSerHisValValAsnThr 680

Db       2005 TCCTGTCAAAACAAGGTGTATTCAAATCTGGAAGGCAACAGGTCTCATGTGGTCAACACT 2064

Qy        681 GlyTrpAlaMetLeuAlaLeuIleAspAlaGlyGlnAlaLysArgAspSerGlnProLeu 700

Db       2065 GGGTGGGCTATGTTGGCTCTCATTGATGCTGGACAGGCTAAGAGAGATTCGCAACCACTG 2124

Qy        701 HisArgAlaAlaAlaTyrLeuIleAsnSerGlnLeuGluAspGlyAspProProGlnGln 720

Db       2125 CACCGGGCAGCTGCATACTTGATAAATTCCCAATTGGAGGACGGTGACTTTCCGCAGCAG 2184

Qy        721 GluIleMetGlyValPheAsnLysAsnCysMetIleThrTyrAlaAlaTyrArgAsnIle 740

Db       2185 GAAATAATGGGAGTCTTCAACAAGAATTGCATGATCACATACGCCGCATACAGAAACATA 2244

Qy        741 ProProIleTrpAlaLeuGlyGluTyrGlnSerGlnValLeuGlnSerArg 757

Db       2245 TTCCCCATTTGGGCGTTGGGAGAATACCAATCCCAAGTATTGCAATCTCGT 2295
```

RESULT 2
US-09-938-842A-833
; Sequence 833, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIP1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 833
; LENGTH: 2280
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-833

Alignment Scores:

Pred. No.:	0	Length:	2280
Score:	3480.00	Matches:	607
Percent Similarity:	90.72%	Conservative:	77
Best Local Similarity:	80.50%	Mismatches:	70
Query Match:	83.90%	Indels:	0
DB:	9	Gaps:	0

US-09-419-679-6 (1-757) x US-09-938-842A-833 (1-2280)

```
Qy          1 MetTrpLysLeuLysPheAlaGluGlyGlyAsnProTrpLeuArgThrLeuAsnAsnHis 20

Db          1 ATGTGGAAACTGAAGATCGCGGAAGGAGGTAGTCCATGGCTTAGAACCACCAATAATCAC 60

Qy         21 ValGlyArgGlnValTrpGluPheAspProLysLeuGlySerProGlnAspLeuLeuGlu 40

Db         61 GTCGGAAGACAGTTTTGGGAGTTCGATCCGAATCTCGGTACTCCTGAGGATCTCGCCGCC 120

Qy         41 IleGluLysAlaArgGlnAsnPheHisAspAsnArgPheThrHisLysHisSerAlaAsp 60

Db        121 GTCGAAGAAGCTAGGAAGTCTTTTTCGGATAATCGATTCGTGCAGAAACATAGCGCCGAT 180

Qy         61 LeuLeuMetArgMetGlnPheAlaArgGluAsnProThrArgGluValLeuProLysVal 80

Db        181 CTGCTTATGCGCCTTCAGTTTTCAAGAGAAAATTTGATTAGCCCAGTTTTACCTCAAGTC 240

Qy         81 GlyValLysAspIleGluAspValThrGlnGluIleValThrLysThrLeuArgArgAla 100

Db        241 AAAATCGAAGACACTGATGATGTTACAGAGGAGATGGTGGAAACCACGTTAAAGAGGGGT 300

Qy        101 ValSerPheHisSerThrLeuGlnCysHisAspGlyHisTrpProGlyAspTyrGlyGly 120

Db        301 CTAGATTTCTATTCAACTATACAGGCACACGACGGGCACTGGCCAGGTGATTATGGTGGT 360

Qy        121 ProMetPheLeuMetProGlyLeuValIleThrLeuSerIleThrGlyAlaLeuAsnThr 140

Db        361 CCTATGTTTCTTCTCCCAGGACTGATAATTACACTCTCCATAACTGGAGCACTGAATACA 420

Qy        141 ValLeuThrGluGluHisArgLysGluIleCysArgTyrLeuTyrAsnHisGlnAsnLys 160

Db        421 GTATTGTCGGAACAACATAAACAAGAAATGCGCCGTTATCTCTATAATCACCAGAATGAG 480

Qy        161 AspGlyGlyTrpGlyLeuHisIleGluGlyProSerThrMetPheGlySerValLeuSer 180

Db        481 GACGGAGGTTGGGGTTTACATATTGAGGGCCCTAGCACCATGTTTGGGTCTGTGTTGAAC 540

Qy        181 TyrIleThrLeuArgLeuLeuGlyGluGlyProAsnAspGlyGlnGlyGluMetGluLys 200

Db        541 TATGTTACTCTAAGGTTGCTTGGAGAAGGACCTAACGATGGAGATGGAGATATGGAGAAA 600

Qy        201 AlaArgAspTrpIleLeuGlyHisGlyGlyAlaThrTyrIleThrSerTrpGlyLysMet 220

Db        601 GGACGAGACTGGATACTAAATCATGGTGGTGCTACCAATATTACATCTTGGGGGAAAATG 660

Qy        221 TrpLeuSerValLeuGlyValTyrGluTrpSerGlyAsnAsnProLeuProProGluIle 240

Db        661 TGGCTATCGGTACTTGGAGCTTTTGAATGGTCCGGAAATAACCCACTGCCACCTGAGATA 720

Qy        241 TrpLeuLeuProTyrMetLeuProPheHisProGlyArgMetTrpCysHisCysArgMet 260

Db        721 TGGCTTCTCCCATATTTCCTGCCAATACATCCAGGAAGGATGTGGTGCCATTGTCGAATG 780

Qy        261 ValTyrLeuProMetSerTyrLeuTyrGlyLysArgPheValGlyProIleSerProThr 280

Db        781 GTGTACTTGCCGATGTCGTATTTGTATGGAAAAAGGTTTGTGGGTCCCATAACGTCCACT 840

Qy        281 ValLeuSerLeuArgLysGluLeuTyrThrValProTyrHisAspIleAspTrpAspGln 300

Db        841 GTTTTATCACTGAGAAAGGAGCTTTTCACACTACCATATCATGAAGTCAACTGGAATGAA 900

Qy        301 AlaArgAsnLeuCysAlaLysGluAspLeuTyrTyrProHisProLeuValGlnAspIle 320

Db        901 GCACGCAACCTTTGCGCAAAGGAGGATTTATACTACCCACATCCACTTGTGCAAGATATT 960

Qy        321 LeuTrpAlaSerLeuHisLysPheLeuGluProIleLeuMetHisTrpProGlyLysArg 340

Db        961 CTTTGGGCATCACTTCATAAGATTGTTGAGCCTGTTCTGATGCGATGGCCTGGTGCAAAT 1020

Qy        341 LeuArgGluLysAlaIleIleSerAlaLeuGluHisIleHisTyrGluAspGluAsnThr 360

Db       1021 TTGAGAGAAAAGGCTATAAGAACCGCAATAGAACATATTCATTATGAAGATGAGAATACT 1080

Qy        361 ArgTyrIleCysIleGlyProValAsnLysValLeuAsnMetLeuCysCysTrpValGlu 380

Db       1081 AGGTACATCTGCATAGGTCCCGTGAACAAGGTATTAAATATGCTTTGCTGTTGGGTAGAA 1140

Qy        381 AspProAsnSerGluAlaPheLysLeuHisLeuProArgIleTyrAspTyrLeuTrpIle 400

Db       1141 GACCCAAACTCAGAGGCTTTCAAGTTGCACCTACCAAGAATCCATGACTTTCTCTGGTTA 1200
```

```
;  CURRENT APPLICATION NUMBER: US/10/607,726
;  CURRENT FILING DATE:  2003-06-27
;  PRIOR APPLICATION NUMBER: US/09/419,679
;  PRIOR FILING DATE: 1999-10-15
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/105,405
;  PRIOR FILING DATE: EARLIER FILING DATE: 1998-10-23
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 5
;   LENGTH: 2560
;   TYPE: DNA
;   ORGANISM: Glycine max
US-10-607-726-5

Alignment Scores:
Pred. No.:                0              Length:          2560
Score:                    4148.00        Matches:         757
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       20             Gaps:            0

US-09-419-679-6 (1-757) x US-10-607-726-5 (1-2560)

Qy          1 MetTrpLysLeuLysPheAlaGluGlyGlyAsnProTrpLeuArgThrLeuAsnAsnHis 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         25 ATGTGGAAGCTCAAGTTCGCCGAAGGAGGGAATCCATGGCTTCGGACATTGAACAATCAC 84

Qy         21 ValGlyArgGlnValTrpGluPheAspProLysLeuGlySerProGlnAspLeuLeuGlu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         85 GTTGGAAGACAGGTGTGGGAGTTCGATCCTAAGCTTGGATCGCCGCAAGATCTCCTCGAG 144

Qy         41 IleGluLysAlaArgGlnAsnPheHisAspAsnArgPheThrHisLysHisSerAlaAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        145 ATTGAGAAAGCTCGCCAGAATTTTCACGATAACCGCTTTACCCACAAACACAGCGCTGAT 204

Qy         61 LeuLeuMetArgMetGlnPheAlaArgGluAsnProThrArgGluValLeuProLysVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        205 CTACTTATGCGGATGCAGTTCGCAAGAGAGAACCCAACACGTGAAGTCTTGCCCAAAGTC 264

Qy         81 GlyValLysAspIleGluAspValThrGlnGluIleValThrLysThrLeuArgArgAla 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        265 GGAGTTAAGGATATTGAGGATGTGACCCAAGAGATTGTGACAAAAACATTAAGAAGGGCC 324

Qy        101 ValSerPheHisSerThrLeuGlnCysHisAspGlyHisTrpProGlyAspTyrGlyGly 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        325 GTAAGTTTCCATTCAACTCTCCAGTGCCATGACGGACACTGGCCGGGAGATTATGGAGGT 384

Qy        121 ProMetPheLeuMetProGlyLeuValIleThrLeuSerIleThrGlyAlaLeuAsnThr 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        385 CCCATGTTTCTGATGCCTGGCTTGGTAATTACTCTGTCTATCACTGGGGCGTTGAATACA 444

Qy        141 ValLeuThrGluGluHisArgLysGluIleCysArgTyrLeuTyrAsnHisGlnAsnLys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        445 GTCTTAACTGAAGAACATAGAAAGGAAATATGCCGTTACCTCTATAATCATCAAAACAAG 504

Qy        161 AspGlyGlyTrpGlyLeuHisIleGluGlyProSerThrMetPheGlySerValLeuSer 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        505 GATGGTGGGTGGGGTTTGCATATTGAAGGTCCAAGCACCATGTTTGGCTCTGTCTTGAGT 564

Qy        181 TyrIleThrLeuArgLeuLeuGlyGluGlyProAsnAspGlyGlnGlyGluMetGluLys 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        565 TATATTACTCTGAGATTGCTAGGTGAGGGGCCTAATGATGGACAAGGGGAAATGGAGAAG 624

Qy        201 AlaArgAspTrpIleLeuGlyHisGlyGlyAlaThrTyrIleThrSerTrpGlyLysMet 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        625 GCACGTGACTGGATTCTAGGGCATGGTGGTGCTACTTATATAACGTCATGGGGGAAGATG 684

Qy        221 TrpLeuSerValLeuGlyValTyrGluTrpSerGlyAsnAsnProLeuProProGluIle 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        685 TGGCTTTCAGTACTTGGAGTGTATGAATGGTCTGGAAATAATCCCCTGCCCCCTGAGATA 744

Qy        241 TrpLeuLeuProTyrMetLeuProPheHisProGlyArgMetTrpCysHisCysArgMet 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        745 TGGCTCCTTCCATACATGCTTCCATTTCATCCAGGAAGGATGTGGTGTCACTGCCGGATG 804

Qy        261 ValTyrLeuProMetSerTyrLeuTyrGlyLysArgPheValGlyProIleSerProThr 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        805 GTCTATTTGCCGATGTCCTACTTATATGGCAAGAGGTTTGTTGGTCCAATCTCACCAACA 864

Qy        281 ValLeuSerLeuArgLysGluLeuTyrThrValProTyrHisAspIleAspTrpAspGln 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        865 GTATTATCTTTGAGAAAAGAGCTTTATACAGTACCATACCATGATATAGATTGGGATCAG 924

Qy        301 AlaArgAsnLeuCysAlaLysGluAspLeuTyrTyrProHisProLeuValGlnAspIle 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        925 GCTCGCAATTTGTGTGCAAAGGAAGATTTGTACTATCCTCACCCACTTGTACAGGATATT 984

Qy        321 LeuTrpAlaSerLeuHisLysPheLeuGluProIleLeuMetHisTrpProGlyLysArg 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        985 CTTTGGGCATCTCTACACAAGTTCCTTGAGCCTATTCTGATGCATTGGCCTGGAAAAAGA 1044

Qy        341 LeuArgGluLysAlaIleIleSerAlaLeuGluHisIleHisTyrGluAspGluAsnThr 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1045 TTGAGGGAAAAGGCTATTATTTCTGCATTGGAGCATATACATTACGAAGATGAGAATACT 1104

Qy        361 ArgTyrIleCysIleGlyProValAsnLysValLeuAsnMetLeuCysCysTrpValGlu 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1105 CGATATATTTGCATAGGTCCTGTAAATAAGGTGTTAAATATGCTTTGCTGTTGGGTGGAA 1164

Qy        381 AspProAsnSerGluAlaPheLysLeuHisLeuProArgIleTyrAspTyrLeuTrpIle 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1165 GATCCAAATTCTGAGGCCTTCAAGTTGCATCTTCCCAGGATTTATGATTATCTATGGATT 1224

Qy        401 AlaGluAspGlyMetLysMetGlnGlyTyrAsnGlySerGlnLeuTrpAspThrAlaPhe 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1225 GCAGAAGATGGCATGAAAATGCAGGGCTACAATGGAAGTCAACTATGGGACACTGCTTTT 1284

Qy        421 AlaValGlnAlaIleIleAlaSerAsnLeuIleGluGluPheGlyProThrIleArgLys 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1285 GCTGTCCAAGCAATTATTGCATCTAACCTCATTGAAGAATTTGGTCCAACTATAAGAAAA 1344

Qy        441 AlaHisThrTyrIleLysAsnSerGlnValLeuGluAspCysProGlyAspLeuAsnLys 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1345 GCTCATACCTATATTAAGAATTCACAGGTTTTAGAAGATTGTCCAGGTGATCTTAATAAA 1404

Qy        461 TrpTyrArgHisIleSerLysGlyAlaTrpProPheSerThrGlyAspHisGlyTrpPro 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1405 TGGTACCGTCACATTTCAAAAGGTGCTTGGCCTTTTTCAACTGGAGATCATGGATGGCCA 1464

Qy        481 IleSerAspCysThrAlaGluGlyLeuLysAlaValLeuLeuLeuSerLysIleAlaPro 500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1465 ATTTCTGACTGCACAGCTGAAGGACTGAAAGCTGTTCTATTACTATCCAAAATTGCACCA 1524

Qy        501 GluIleValGlyGluProIleAspValLysArgLeuTyrAspSerValAsnValIleLeu 520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1525 GAAATAGTTGGTGAGCCAATAGACGTGAAGCGATTATATGATTCTGTAAATGTCATTCTC 1584

Qy        521 SerLeuGlnAsnGluAspGlyGlyPheAlaThrTyrGluLeuLysArgSerTyrAsnTrp 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1585 TCACTACAGAATGAAGATGGTGGTTTTGCAACATATGAGCTTAAACGATCTTATAATTGG 1644

Qy        541 LeuGluIleIleAsnProAlaGluThrPheGlyAspIleValIleAspTyrProTyrVal 560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1645 TTGGAGATAATCAATCCTGCTGAAACTTTTGGTGACATCGTTATTGATTATCCTTATGTG 1704

Qy        561 GluCysThrSerAlaAlaIleGlnAlaLeuAlaSerPheArgLysLeuTyrProGlyHis 580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1705 GAATGTACATCAGCAGCGATTCAAGCTTTGGCATCATTTAGGAAATTATATCCTGGGCAT 1764

Qy        581 ArgArgGluGluIleGlnHisCysIleAspLysAlaThrThrPheIleGluLysIleGln 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1765 CGCCGAGAAGAAATACAACATTGTATCGATAAAGCCACTACCTTCATTGAAAAAATACAA 1824

Qy        601 AlaSerAspGlySerTrpTyrGlySerTrpGlyValCysPheThrTyrGlyAlaTrpPhe 620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 5, 2005, 12:49:46 ; Search time 1460 Seconds
(without alignments)
4287.845 Million cell updates/sec

Title: US-09-419-679-6
Perfect score: 4148
Sequence: 1 MWKLKFAEGGNPWLRTLNNH.....RNIFPIWALGEYQSQVLQSR 757

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 9794790 seqs, 4134909567 residues

Total number of hits satisfying chosen parameters: 19589580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/6630617/runat_02112005_120302_15132/app_query.fasta_1.903
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=6630617_@CGN_1_1_512_@runat_02112005_120302_15132
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq:*
20: /cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq:*
21: /cgn2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq:*
22: /cgn2_6/ptodata/2/pubpna/US10I_PUBCOMB.seq:*
23: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
24: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq2:*
25: /cgn2_6/ptodata/2/pubpna/US11A_PUBCOMB.seq:*
26: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
27: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
28: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4148	100.0	2560	20	US-10-607-726-5	Sequence 5, Appli
2	3480	83.9	2280	9	US-09-938-842A-833	Sequence 833, App
3	3480	83.9	2280	11	US-09-938-842A-833	Sequence 833, App
4	3248	78.3	2776	21	US-10-425-115-91064	Sequence 91064, A
5	3240.5	78.1	4120	20	US-10-437-963-71137	Sequence 71137, A
6	3237	78.0	2638	19	US-10-425-114-25096	Sequence 25096, A
7	3237	78.0	2786	21	US-10-425-115-90182	Sequence 90182, A
8	3234	78.0	2355	20	US-10-437-963-71159	Sequence 71159, A
9	3181.5	76.7	2653	19	US-10-425-114-24593	Sequence 24593, A
10	3125	75.3	2558	20	US-10-607-726-1	Sequence 1, Appli
11	2551	61.5	2538	18	US-10-221-074-7	Sequence 7, Appli
12	2543.5	61.3	2593	19	US-10-425-114-24645	Sequence 24645, A
13	2543.5	61.3	2627	21	US-10-425-115-65768	Sequence 65768, A
14	2541	61.3	2669	19	US-10-424-599-35539	Sequence 35539, A
15	2538	61.2	2478	18	US-10-221-074-3	Sequence 3, Appli
16	2538	61.2	2478	18	US-10-427-570A-4	Sequence 4, Appli
17	2538	61.2	2567	19	US-10-425-114-7892	Sequence 7892, Ap
18	2522.5	60.8	2574	21	US-10-739-930-2894	Sequence 2894, Ap
19	2516.5	60.7	2093	18	US-10-259-194A-115	Sequence 115, App
20	2454.5	59.2	2731	18	US-10-221-074-1	Sequence 1, Appli
21	2436	58.7	2746	19	US-10-424-599-31007	Sequence 31007, A
22	2436	58.7	2766	18	US-10-221-074-5	Sequence 5, Appli
23	2436	58.7	2766	18	US-10-427-570A-1	Sequence 1, Appli
24	2433	58.7	1882	20	US-10-607-726-3	Sequence 3, Appli
25	2363.5	57.0	2416	19	US-10-424-599-35538	Sequence 35538, A
26	2354.5	56.8	2457	17	US-10-168-445-14	Sequence 14, Appl
27	2339.5	56.4	2319	19	US-10-425-114-1854	Sequence 1854, Ap
28	2339.5	56.4	2379	21	US-10-425-115-90180	Sequence 90180, A
29	2336	56.3	2484	19	US-10-425-114-10317	Sequence 10317, A
30	2336	56.3	2500	19	US-10-424-599-4661	Sequence 4661, Ap
31	2311.5	55.7	2067	18	US-10-429-949-6	Sequence 6, Appli
32	2264	54.6	2510	19	US-10-425-114-24639	Sequence 24639, A
33	2256.5	54.4	2415	21	US-10-739-930-740	Sequence 740, App
34	2244	54.1	2526	21	US-10-425-115-44214	Sequence 44214, A
35	2226.5	53.7	2179	19	US-10-425-114-9157	Sequence 9157, Ap
36	2097.5	50.6	2432	20	US-10-437-963-100852	Sequence 100852,
37	2083	50.2	2639	18	US-10-389-966-248	Sequence 248, App
38	2070.5	49.9	1939	19	US-10-425-114-10363	Sequence 10363, A
39	2060.5	49.7	2182	19	US-10-424-599-139406	Sequence 139406,
40	1959	47.2	2414	21	US-10-425-115-171124	Sequence 171124,
41	1882	45.4	1546	18	US-10-259-194A-548	Sequence 548, App
42	1875	45.2	1862	20	US-10-437-963-63249	Sequence 63249, A
43	1858	44.8	1729	19	US-10-425-114-10855	Sequence 10855, A
44	1829	44.1	1842	19	US-10-425-114-10336	Sequence 10336, A
45	1784	43.0	1869	19	US-10-424-599-53604	Sequence 53604, A

ALIGNMENTS

RESULT 1
US-10-607-726-5
; Sequence 5, Application US/10607726
; Publication No. US20040121439A1
; GENERAL INFORMATION:
; APPLICANT: Falco, S. Carl
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Hitz, William D.
; APPLICANT: Kinney, Anthony J.
; APPLICANT: Rafalski, Antoni
; APPLICANT: McGonigle, Brian
; APPLICANT: Lohman, Karin
; TITLE OF INVENTION: Enzymes Involved In Squalene Metabolizm
; FILE REFERENCE: BB1113 US NA

THIS PAGE BLANK (USPTO)

```
Qy       517 AsnValIleLeuSerLeuGlnAsnGluAspGlyGlyPheAlaThrTyrGluLeuLysArg 536
Db      1558 AATCTCTTGCTATCGTTGCAGGGTGAAAAAGGAGGTTTGACTGCATGGGAGCCTGTCCGT 1617

Qy       537 SerTyrAsnTrpLeuGluIleIleAsnProAlaGluThrPheGlyAspIleValIleAsp 556
Db      1618 GCACAAGAATGGCTGGAATTGCTCAATCCCACAGATTTCTTTACTTGTGTTATGGCTGAA 1677

Qy       557 TyrProTyrValGluCysThrSerAlaAlaIleGlnAlaLeuAlaSerPheArgLysLeu 576
Db      1678 CGCGAGTATGTAGAATGTACCTCAGCTGTTATACAAGCTTTGGTACTGTTCAAACAACTT 1737

Qy       577 TyrProGlyHisArgArgGluGluIleGlnHisCysIleAspLysAlaThrThrPheIle 596
Db      1738 TATCCGGATCACAGGACAAAAGAGATCATCAAGTCGATTGAGAAAGGGGTGCAATTCATT 1797

Qy       597 GluLysIleGlnAlaSerAspGlySerTrpTyrGlySerTrpGlyValCysPheThrTyr 616
Db      1798 GAAAGCAAACAAACGCCTGATGGTTCATGGCATGGAAACTGGGGTATCTGTTTCATCTAC 1857

Qy       617 GlyAlaTrpPheGlyValLysGlyLeuIleAlaAlaGlyArgSerPheSerAsnCysSer 636
Db      1858 GCTACTTGGTTTGCTCTGAGCGGCCTAGCAGCCGCTGGTAAAACTTACAAAAGTTGCCTG 1917

Qy       637 SerIleArgLysAlaCysGluPheLeuLeuSerLysGlnLeuProSerGlyGlyTrpGly 656
Db      1918 GCTGTGCGCAAAGGTGTAGATTTCCTGCTTGCGATACAGGAAGAAGATGGAGGTTGGGGC 1977

Qy       657 GluSerTyrLeuSerCysGlnAsnLysValTyrSerAsnLeuGluGlyAsnArgSerHis 676
Db      1978 GAAAGCCATCTGTCATGCCCTGAGCAGAGATACATACCATTAGAAGGGAACAGATCAAAC 2037

Qy       677 ValValAsnThrGlyTrpAlaMetLeuAlaLeuIleAspAlaGlyGlnAlaLysArgAsp 696
Db      2038 CTAGTGCAAACCGCATGGGCTATGATGGGTTTGATTCATGCCGGACAGGCCGAGAGAGAT 2097

Qy       697 SerGlnProLeuHisArgAlaAlaAlaTyrLeuIleAsnSerGlnLeuGluAspGlyAsp 716
Db      2098 CCTACACCTCTTCACCGTGCTGCGAAACTTATCATCACTTCACAACTTGAAAATGGGGAC 2157

Qy       717 PheProGlnGlnGluIleMetGlyValPheAsnLysAsnCysMetIleThrTyrAlaAla 736
Db      2158 TTTCCGCAACAGGAAATATTAGGAGTGTTCATGAATACATGCATGCTACACTATGCTACG 2217

Qy       737 TyrArgAsnIlePheProIleTrpAlaLeuGlyGluTyrGln 750
Db      2218 TACAGAAACATCTTCCCACTATGGGCACTCGCGGAATATCGG 2259
```

Search completed: November 5, 2005, 13:07:31
Job time : 1078 secs

PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Alignment Scores:

Pred. No.:	4.1e-228	Length:	2532
Score:	2418.50	Matches:	435
Percent Similarity:	73.47%	Conservative:	119
Best Local Similarity:	57.69%	Mismatches:	195
Query Match:	58.31%	Indels:	5
DB:	3	Gaps:	4

US-09-419-679-6 (1-757) x AAC46759 (1-2532)

```
Qy          1 MetTrpLysLeuLysPheAlaGluGlyGly------AsnProTrpLeuArgThrLeuAsn 18
              |||||||||||||||||      ||||||         :::|||:::|||   :::   |||
Db          1 ATGTGGAAGTTGAAGATAGGAGAGGGAAATGGAGAAGATCCTTACTTGTTCAGCAGCAAC 60

Qy         19 AsnHisValGlyArgGlnValTrpGluPheAspProLysLeuGlySerProGlnAspLeu 38
              |||   ||||||||||||   ||||||||||||||||||   |||:::|||::::::
Db         61 AACTTCGTCGGACGTCAAACATGGGAGTTTGATCCCAAAGCCGGCACACCTGAGGAACGA 120

Qy         39 LeuGluIleGluLysAlaArgGlnAsnPheHisAspAsnArgPheThrHisLysHisSer 58
                 :::|||   ||||||:::|||:::   |||||||||          |||
Db        121 GCCGCCGTCGAAGATGCTCGCCGGAACTATTTAGACAACCGTCCCCGTGTTAAAGGTTGC 180

Qy         59 AlaAspLeuLeuMetArgMetGlnPheAlaArgGluAsnProThrArgGluValLeuPro 78
              :::|||||||||   ||||||||||||   :::|||          :::|||:::|||
Db        181 AGTGATCTCTTGTGGCGAATGCAATTTTTGAAAGAGGCGAAATTCGAGCAAGTGATCCCG 240

Qy         79 LysValGlyValLysAspIleGluAspValThrGlnGluIleValThrLysThrLeuArg 98
                 |||   :::   |||   |||   :::|||   :::       |||       ||||||
Db        241 CCGGTGAAGATCGACGACGGCGAAGGCATAACTTACAAAAACGCGACGGATGCGTTACGG 300

Qy         99 ArgAlaValSerPheHisSerThrLeuGlnCysHisAspGlyHisTrpProGlyAspTyr 118
              ||||||||||||||:::|||   ||||||      ||||||||||||||   :::
Db        301 AGAGCAGTTTCTTTCTACTCGGCTTTGCAGTCCTCCGATGGCCACTGGCCGGCGGAAATC 360

Qy        119 GlyGlyProMetPheLeuMetProGlyLeuValIleThrLeuSerIleThrGlyAlaLeu 138
                 |||   :::|||   :::|||   ||||||          |||||||||   |||
Db        361 ACCGGAACTCTCTTCTTCCTTCCTCCATTGGTATTTTGTTTCTATATCACAGGACACCTC 420

Qy        139 AsnThrValLeuThrGluGluHisArgLysGluIleCysArgTyrLeuTyrAsnHisGln 158
                 :::          ||||||||||||||:::   |||:::::|||   ||||||
Db        421 GAGAAGATATTTGATGCAGAACATCGCAAAGAGATGCTTCGGCATATCTATTGCCACCAG 480

Qy        159 AsnLysAspGlyGlyTrpGlyLeuHisIleGluGlyProSerThrMetPheGlySerVal 178
              |||:::|||||||||||||||||||||||||||||||   |||   ||||||   :::|||
Db        481 AACGAAGACGGTGGATGGGGTTTACATATTGAGGGAAAGAGCGTTATGTTCTGCACCGTA 540

Qy        179 LeuSerTyrIleThrLeuArgLeuLeuGlyGluGlyProAsnAspGlyGlnGlyGlu--- 197
              |||:::||||||   ||||||:::|||||||||||||||||||   |||:::
Db        541 CTGAATTACATATGCTTGCGTATGCTCGGAGAAGGTCCCAATGGAGGGCGCAATAACGCC 600

Qy        198 MetGluLysAlaArgAspTrpIleLeuGlyHisGlyGlyAlaThrTyrIleThrSerTrp 217
                 ::::::||||||   |||||||||   |||||||||   |||||||||   ||||||
Db        601 TGCAAACGGGCCAGGCAATGGATTCTTGACCATGGTGGTGTGACTTATATTCCTTCTTGG 660

Qy        218 GlyLysMetTrpLeuSerValLeuGlyValTyrGluTrpSerGlyAsnAsnProLeuPro 237
              ||||||:::|||||||||:::||||||:::|||:::|||||||||   ||||||:::|||
Db        661 GGAAAAATTTGGCTCTCGATACTCGGAATCTATGATTGGTCTGGAACCAACCCAATGCCT 720

Qy        238 ProGluIleTrpLeuLeuProTyrMetLeuProPheHisProGlyArgMetTrpCysHis 257
              ||||||||||||||||||||          |||   |||   |||:::       |||:::
Db        721 CCCGAGATTTGGTTGCTGCCTTCTTTCTTTCCAATACACTTAGGAAAAACTTTGTGTTAT 780

Qy        258 CysArgMetValTyrLeuProMetSerTyrLeuTyrGlyLysArgPheValGlyProIle 277
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||:::
Db        781 ACCCGGATGGTTTATATGCCCATGTCTTATCTATATGGGAAACGATTTGTTGGTCCTCTT 840

Qy        278 SerProThrValLeuSerLeuArgLysGluLeuTyrThrValProTyrHisAspIleAsp 297
              :::|||   ::::::   ||||||||||||||:::       ||||||   :::|||:::
Db        841 ACACCTCTTATTATGCTATTGCGCAAAGAACTGCACTTGCAACCTTATGAGGAAATCAAT 900

Qy        298 TrpAspGlnAlaArgAsnLeuCysAlaLysGluAspLeuTyrTyrProHisProLeuVal 317
              |||::::::||||||   ||||||||||||||||:::   ||||||||||||||||||
Db        901 TGGAACAAAGCGCGCCGTCTATGTGCAAAAGAAGACATGATTTATCCTCATCCTCTGGTT 960

Qy        318 GlnAspIleLeuTrpAlaSerLeuHisLysPheLeuGluProIleLeuMetHisTrpPro 337
              ||||||:::|||||||   :::||||||   |||:::||||||||||||   :::||||||
Db        961 CAAGATTTGTTATGGGACACTCTTCACAATTTTGTGGAGCCTATCCTTACAAATTGGCCG 1020

Qy        338 GlyLysArgLeu---ArgGluLysAlaIleIleSerAlaLeuGluHisIleHisTyrGlu 356
                 |||:::|||   ||||||||||||:::      |||:::||||||||||||||||||
Db       1021 TTAAAAAAACTTGTACGGGAAAAGGCTCTTCGAGTGGCAATGGAACACATACATTATGAG 1080

Qy        357 AspGluAsnThrArgTyrIleCysIleGlyProValAsnLysValLeuAsnMetLeuCys 376
              |||||||||:::   ||||||   ||||||   |||   |||||||||   ||||||
Db       1081 GACGAAAATAGCCATTATATTACCATCGGATGTGTTGAGAAGGTTCTGTGCATGCTTGCT 1140

Qy        377 CysTrpValGluAspProAsnSerGluAlaPheLysLeuHisLeuProArgIleTyrAsp 396
              ||||||:::|||:::||||||   :::   ||||||   ||||||   ||||||   |||
Db       1141 TGCTGGATCGAGAATCCTAATGGAGATCACTTTAAGAAACATCTCGCTAGAATTCCGGAC 1200

Qy        397 TyrLeuTrpIleAlaGluAspGlyMetLysMetGlnGlyTyrAsnGlySerGlnLeuTrp 416
              ::::::|||:::||||||||||||:::|||||||||   :::   ||||||||||||||||
Db       1201 TTCATGTGGGTTGCTGAAGACGGACTGAAAATGCAGAGCTTT---GGAAGTCAACTCTGG 1257

Qy        417 AspThrAlaPheAlaValGlnAlaIleIleAlaSerAsnLeuIleGluGluPheGlyPro 436
              ||||||   ||||||:::||||||:::|||   :::|||   :::|||
Db       1258 GACACAGTTTTTGCGATTCAAGCTTTACTTGCTTGTGACCTTTCTGACGAAACTGATGAT 1317

Qy        437 ThrIleArgLysAlaHisThrTyrIleLysAsnSerGlnValLeuGluAspCysProGly 456
              :::||||||   |||:::::||||||   |||||||||   |||:::       |||
Db       1318 GTACTCAGGAAAGGACACAGTTTCATCAAAAAATCCCAGGTTAGAGAAAACCCTTCAGGT 1377

Qy        457 AspLeuAsnLysTrpTyrArgHisIleSerLysGlyAlaTrpProPheSerThrGlyAsp 476
              |||                 ||||||||||||||||||||||||||||   |||   |||
Db       1378 GACTTTAAGAGCATGTATCGCCATATTTCCAAAGGAGCCTGGACTCTGTCCGATCGAGAT 1437

Qy        477 HisGlyTrpProIleSerAspCysThrAlaGluGlyLeuLysAlaValLeuLeuLeuSer 496
              |||||||||   :::||||||||||||||||||   ||||||      :::|||||||||
Db       1438 CATGGATGGCAAGTCTCAGATTGTACAGCTGAAGCTTTGAAGTGTTGCATGCTGCTCTCC 1497

Qy        497 LysIleAlaProGluIleValGlyGluProIleAspValLysArgLeuTyrAspSerVal 516
                 :::      |||:::||||||:::   ||||||   ::::::||||||||||||||||
Db       1498 ATGATGCCAGCTGAGGTCGTTGGCCAGAAAATAGATCCTGAACAATTATATGATTCTGTT 1557
```

PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.

```
XX
SQ   Sequence 1882 BP; 565 A; 363 C; 441 G; 513 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                9.73e-230      Length:          1882
Score:                    2433.00        Matches:         419
Percent Similarity:       90.43%         Conservative:    63
Best Local Similarity:    78.61%         Mismatches:      51
Query Match:              58.65%         Indels:          0
DB:                       12             Gaps:            0

US-09-419-679-6 (1-757) x ADQ95185 (1-1882)

Qy        222 LeuSerValLeuGlyValTyrGluTrpSerGlyAsnAsnProLeuProProGluIleTrp 241
              ||||||||||||||||||:::::|||||||||||||||||||:::|||||||||||||||||
Db         15 CTTTCGGTGCTTGGTGTATTTGACTGGTCTGGCAACAACCCAGTGCCACCAGAAATATGG 74

Qy        242 LeuLeuProTyrMetLeuProPheHisProGlyArgMetTrpCysHisCysArgMetVal 261
              ||||||||||||   ||||||  ||||||||||||||||||||||||||||||||||||||
Db         75 TTGTTGCCATATTTCCTGCCGATTCATCCAGGGCGAATGTGGTGTCATTGCCGGATGGTT 134

Qy        262 TyrLeuProMetSerTyrLeuTyrGlyLysArgPheValGlyProIleSerProThrVal 281
              ||||||||||||   |||:::|||||||||||||||||||||||:::::|||   :::
Db        135 TATTTGCCTATGTGTTACATTTATGGAAAGAGGTTTGTGGGCCCAGTTACACCAATTATA 194

Qy        282 LeuSerLeuArgLysGluLeuTyrThrValProTyrHisAspIleAspTrpAspGlnAla 301
              |||   ||||||||||||||||   |||||||||:::::::::|||||||||:::|||
Db        195 TTGGAATTAAGAAAGGAACTCTACGAAGTACCCTACAATGAAGTTGATTGGGACAAGGCT 254

Qy        302 ArgAsnLeuCysAlaLysGluAspLeuTyrTyrProHisProLeuValGlnAspIleLeu 321
              ||||||||||||||||||||||||||||||||||||||||   |||||||||:::|||
Db        255 CGCAATCTATGTGCTAAGGAAGATCTGTACTATCCACATCCATTCGTGCAGGATGTATTA 314

Qy        322 TrpAlaSerLeuHisLysPheLeuGluProIleLeuMetHisTrpProGlyLysArgLeu 341
              ||||||:::||||||||||||:::||||||   ::::::   |||||||||   :::|||
Db        315 TGGGCCACTCTCCACAAATTTGTTGAACCAGCTATGTTGCGTTGGCCTGGGAACAAATTG 374

Qy        342 ArgGluLysAlaIleIleSerAlaLeuGluHisIleHisTyrGluAspGluAsnThrArg 361
              ||||||||||||:::   :::   ::::::||||||||||||||||||||||||||||||
Db        375 AGGGAGAAAGCTTTGGACACTGTCATGCAGCATATTCATTATGAAGATGAGAACACCCGA 434

Qy        362 TyrIleCysIleGlyProValAsnLysValLeuAsnMetLeuCysCysTrpValGluAsp 381
              ||||||||||||||||||||||||||||||||||||||||   ||||||:::||||||
Db        435 TATATTTGCATTGGTCCAGTAAACAAGGTATTAAATATGCTTGCTTGCTGGATTGAAGAT 494

Qy        382 ProAsnSerGluAlaPheLysLeuHisLeuProArgIleTyrAspTyrLeuTrpIleAla 401
              ||||||||||||||||||||||||:::||||||:::::|||||||||||||||||||||
Db        495 CCAAACTCAGAGGCATTCAAACTCCACATTCCAAGAGTCCACGATTACCTATGGATTGCA 554

Qy        402 GluAspGlyMetLysMetGlnGlyTyrAsnGlySerGlnLeuTrpAspThrAlaPheAla 421
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        555 GAAGATGGCATGAAAATGCAGGGTTATAATGGAAGCCAGCTGTGGGACACAGCTTTCACA 614

Qy        422 ValGlnAlaIleIleAlaSerAsnLeuIleGluGluPheGlyProThrIleArgLysAla 441
              ||||||||||||:::|||:::   ||||||||||||||||||||::::::   |||
Db        615 GTTCAAGCTATAGTGGCTACTGGCCTCATTGAAGAATTTGGTCCTACTCTTAAACTAGCA 674

Qy        442 HisThrTyrIleLysAsnSerGlnValLeuGluAspCysProGlyAspLeuAsnLysTrp 461
              |||   |||||||||   :::||||||:::::||||||||||||||||||:::::|||
Db        675 CATGGCTACATAAAGAAAACGCAGGTTATCGATGACTGCCCTGGAGATCTTAGTCAGTGG 734

Qy        462 TyrArgHisIleSerLysGlyAlaTrpProPheSerThrGlyAspHisGlyTrpProIle 481
              ||||||||||||||||||||||||||||||||||   ||||||||||||||||||||||
Db        735 TACCGCCACATATCTAAAGGTGCATGGCCCTTTTCTACTGCTGATCATGGTTGGCCTATA 794

Qy        482 SerAspCysThrAlaGluGlyLeuLysAlaValLeuLeuLeuSerLysIleAlaProGlu 501
              ||||||||||||||||||||||||||||   ||||||||||||||||:::|||:::
Db        795 TCAGATTGCACTGCAGAAGGACTTAAGGCGGCATTATTGCTATCGAAGATTTCTCCAGAT 854

Qy        502 IleValGlyGluProIleAspValLysArgLeuTyrAspSerValAsnValIleLeuSer 521
              ||||||||||||   :::::::|||   ||||||||||||||||||   :::::::|||
Db        855 ATTGTTGGCGAAGCAGTGGAAGTTAATAGACTGTATGATTCTGTCAATTGTTTGATGTCA 914

Qy        522 LeuGlnAsnGluAspGlyGlyPheAlaThrTyrGluLeuLysArgSerTyrAsnTrpLeu 541
              |||:::::|||||||||||||||||||||||||   |||||||||   ||||||
Db        915 TACATGAATGATAATGGTGGATTTGCAACATATGAACTCACAAGGTCTTATGCCTGGCTG 974

Qy        542 GluIleIleAsnProAlaGluThrPheGlyAspIleValIleAspTyrProTyrValGlu 561
              |||:::|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        975 GAGCTTATCAATCCTGCTGAGACCTTTGGGGACATTGTGATTGATTATCCTTATGTGGAA 1034

Qy        562 CysThrSerAlaAlaIleGlnAlaLeuAlaSerPheArgLysLeuTyrProGlyHisArg 581
              ||||||||||||||||||||||||||||||   :::|||:::|||||||||||||||||||
Db       1035 TGCACTTCAGCAGCAATTCAGGCCCTGACAGCATTTAAAAAGCTCTACCCTGGACACCGC 1094

Qy        582 ArgGluGluIleGlnHisCysIleAspLysAlaThrThrPheIleGluLysIleGlnAla 601
              :::   ||||||   :::||||||   ||||||   :::|||||||||   ||||||
Db       1095 AAGAGTGAAATAGACAACTGTATAAGCAAAGCTGCTAGCTTTATTGAGGGTATTCAAAAA 1154

Qy        602 SerAspGlySerTrpTyrGlySerTrpGlyValCysPheThrTyrGlyAlaTrpPheGly 621
              ||||||||||||||||||||||||||||   ||||||||||||||||||   |||||||||
Db       1155 AGCGATGGTTCATGGTATGGTTCTTGGGCTGTTTGTTTTACCTATGGCACATGGTTTGGT 1214

Qy        622 ValLysGlyLeuIleAlaAlaGlyArgSerPheSerAsnCysSerSerIleArgLysAla 641
              ||||||||||||:::||||||||||:::|||   |||      :::||||||||||||
Db       1215 GTAAAGGGATTAGTTGCTGCTGGTAGGACATTCAAAAACAGTCCTGCAATCAGAAAGGCA 1274

Qy        642 CysGluPheLeuLeuSerLysGlnLeuProSerGlyGlyTrpGlyGluSerTyrLeuSer 661
              |||:::||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db       1275 TGTGACTTTTTGTTGTCAAAAGAGCTTCCTTCTGGAGGCTGGGGAGAAAGCTATTTGTCA 1334

Qy        662 CysGlnAsnLysValTyrSerAsnLeuGluGlyAsnArgSerHisValValAsnThrGly 681
              |||::::::||||||:::||||||||||||   |||   |||   ||||||||||||
Db       1335 TCCCAAGATCAGGTTTATACCAATCTCGAAGGGAAGCGACCTCATGCGGTGAACACTGGT 1394

Qy        682 TrpAlaMetLeuAlaLeuIleAspAlaGlyGlnAlaLysArgAspSerGlnProLeuHis 701
              ||||||||||||||||||||||||||||||||||||||::::||||||      |||||||||
Db       1395 TGGGCCATGCTAGCCCTAATCGATGCAGGGCAGGCTGAGAGAGATCCAATTCCTTTGCAT 1454

Qy        702 ArgAlaAlaAlaTyrLeuIleAsnSerGlnLeuGluAspGlyAspPheProGlnGlnGlu 721
              |||||||||      |||||||||   |||   |||||||||:::||||||||||||||||
Db       1455 CGAGCAGCGAAGGTTTTGATCAACTTACAATCGGAAGATGGTGAATTTCCCCAGCAAGAG 1514

Qy        722 IleMetGlyValPheAsnLysAsnCysMetIleThrTyrAlaAlaTyrArgAsnIlePhe 741
              |||:::|||||||||||||||||||||||||||:::|||:::   ||||||||||||||||
Db       1515 ATCATTGGAGTCTTCAACAAAAACTGCATGATCAGCTACTCCGAGTATAGAAACATCTTC 1574

Qy        742 ProIleTrpAlaLeuGlyGluTyrGlnSerGlnValLeu 754
              |||||||||||||||||||||||:::   :::||||||
Db       1575 CCTATTTGGGCCCTTGGGGAGTACCGTCGCCGCGTCTTG 1613

RESULT 15
AAC46759
ID   AAC46759 standard; DNA; 2532 BP.
XX
AC   AAC46759;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 51296.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway; metabolic pathway;
KW   promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-00301439.
XX
PR   25-FEB-1999;   99US-0121825P.
PR   05-MAR-1999;   99US-0123180P.
PR   09-MAR-1999;   99US-0123548P.
```

```
Db        495 CCAAACTCAGAGGCATTCAAACTCCACATTCCAAGAGTCCACGATTACCTATGGATTGCA 554

Qy        402 GluAspGlyMetLysMetGlnGlyTyrAsnGlySerGlnLeuTrpAspThrAlaPheAla 421

Db        555 GAAGATGGCATGAAAATGCAGGGTTATAATGGAAGCCAGCTGTGGGACACAGCTTTCACA 614

Qy        422 ValGlnAlaIleIleAlaSerAsnLeuIleGluGluPheGlyProThrIleArgLysAla 441

Db        615 GTTCAAGCTATAGTGGCTACTGGCCTCATTGAAGAATTTGGTCCTACTCTTAAACTAGCA 674

Qy        442 HisThrTyrIleLysAsnSerGlnValLeuGluAspCysProGlyAspLeuAsnLysTrp 461

Db        675 CATGGCTACATAAAGAAAACGCAGGTTATCGATGACTGCCCTGGAGATCTTAGTCAGTGG 734

Qy        462 TyrArgHisIleSerLysGlyAlaTrpProPheSerThrGlyAspHisGlyTrpProIle 481

Db        735 TACCGCCACATATCTAAAGGTGCATGGCCCTTTTCTACTGCTGATCATGGTTGGCCTATA 794

Qy        482 SerAspCysThrAlaGluGlyLeuLysAlaValLeuLeuLeuSerLysIleAlaProGlu 501

Db        795 TCAGATTGCACTGCAGAAGGACTTAAGGCGGCATTATTGCTATCGAAGATTTCTCCAGAT 854

Qy        502 IleValGlyGluProIleAspValLysArgLeuTyrAspSerValAsnValIleLeuSer 521

Db        855 ATTGTTGGCGAAGCAGTGGAAGTTAATAGACTGTATGATTCTGTCAATTGTTTGATGTCA 914

Qy        522 LeuGlnAsnGluAspGlyGlyPheAlaThrTyrGluLeuLysArgSerTyrAsnTrpLeu 541

Db        915 TACATGAATGATAATGGTGGATTTGCAACATATGAACTCACAAGGTCTTATGCCTGGCTG 974

Qy        542 GluIleIleAsnProAlaGluThrPheGlyAspIleValIleAspTyrProTyrValGlu 561

Db        975 GAGCTTATCAATCCTGCTGAGACCTTTGGGGACATTGTGATTGATTATCCTTATGTGGAA 1034

Qy        562 CysThrSerAlaAlaIleGlnAlaLeuAlaSerPheArgLysLeuTyrProGlyHisArg 581

Db       1035 TGCACTTCAGCAGCAATTCAGGCCCTGACAGCATTTAAAAAGCTCTACCCTGGACACCGC 1094

Qy        582 ArgGluGluIleGlnHisCysIleAspLysAlaThrThrPheIleGluLysIleGlnAla 601

Db       1095 AAGAGTGAAATAGACAACTGTATAAGCAAAGCTGCTAGCTTTATTGAGGGTATTCAAAAA 1154

Qy        602 SerAspGlySerTrpTyrGlySerTrpGlyValCysPheThrTyrGlyAlaTrpPheGly 621

Db       1155 AGCGATGGTTCATGGTATGGTTCTTGGGCTGTTTGTTTTACCTATGGCACATGGTTTGGT 1214

Qy        622 ValLysGlyLeuIleAlaAlaGlyArgSerPheSerAsnCysSerSerIleArgLysAla 641

Db       1215 GTAAAGGGATTAGTTGCTGCTGGTAGGACATTCAAAAACAGTCCTGCAATCAGAAAGGCA 1274

Qy        642 CysGluPheLeuLeuSerLysGlnLeuProSerGlyGlyTrpGlyGluSerTyrLeuSer 661

Db       1275 TGTGACTTTTTGTTGTCAAAAGAGCTTCCTTCTGGAGGCTGGGGAGAAAGCTATTTGTCA 1334

Qy        662 CysGlnAsnLysValTyrSerAsnLeuGluGlyAsnArgSerHisValValAsnThrGly 681

Db       1335 TCCCAAGATCAGGTTTATACCAATCTCGAAGGGAAGCGACCTCATGCGGTGAACACTGGT 1394

Qy        682 TrpAlaMetLeuAlaLeuIleAspAlaGlyGlnAlaLysArgAspSerGlnProLeuHis 701

Db       1395 TGGGCCATGCTAGCCCTAATCGATGCAGGGCAGGCTGAGAGAGATCCAATTCCTTTGCAT 1454

Qy        702 ArgAlaAlaAlaTyrLeuIleAsnSerGlnLeuGluAspGlyAspPheProGlnGlnGlu 721

Db       1455 CGAGCAGCGAAGGTTTTGATCAACTTACAATCGGAAGATGGTGAATTTCCCCAGCAAGAG 1514

Qy        722 IleMetGlyValPheAsnLysAsnCysMetIleThrTyrAlaAlaTyrArgAsnIlePhe 741

Db       1515 ATCATTGGAGTCTTCAACAAAAACTGCATGATCAGCTACTCCGAGTATAGAAACATCTTC 1574

Qy        742 ProIleTrpAlaLeuGlyGluTyrGlnSerGlnValLeu 754

Db       1575 CCTATTTGGGCCCTTGGGGAGTACCGTCGCCGCGTCTTG 1613
```

```
RESULT 14
ADQ95185
ID   ADQ95185 standard; cDNA; 1882 BP.
XX
AC   ADQ95185;
XX
DT   23-SEP-2004  (first entry)
XX
DE   Rice cycloartenol synthase cDNA.
XX
KW   rice; cycloartenol synthase; squalene metabolism; gene; ss.
XX
OS   Oryza sativa.
XX
FH   Key             Location/Qualifiers
FT   CDS             3. .1628
FT                   /*tag=  a
FT                   /product= "Rice cycloartenol synthase"
FT                   /partial
FT                   /note= "No start codon given"
XX
PN   US2004121439-A1.
XX
PD   24-JUN-2004.
XX
PF   27-JUN-2003; 2003US-00607726.
XX
PR   23-OCT-1998;   98US-0105405P.
PR   15-OCT-1999;   99US-00419679.
XX
PA   (FALC/) FALCO S C.
PA   (FAMO/) FAMODU O O.
PA   (HITZ/) HITZ W D.
PA   (KINN/) KINNEY A J.
PA   (RAFA/) RAFALSKI J A.
PA   (MCGO/) MCGONIGLE B.
PA   (LOHM/) LOHMAN K N.
XX
PI   Falco SC,  Famodu OO,  Hitz WD,  Kinney AJ,  Rafalski JA;
PI   Mcgonigle B,  Lohman KN;
XX
DR   WPI; 2004-479684/45.
DR   P-PSDB; ADQ95186.
XX
PT   New composition comprising enzymes involved in squalene metabolism,
PT   useful in altering squalene metabolism in plants and seeds.
XX
PS   Claim 2; SEQ ID NO 3; 33pp; English.
XX
CC   The invention describes a composition comprising an isolated
CC   polynucleotide comprising a nucleotide sequence encoding a first
CC   polypeptide of at least 200 amino acids. A composition comprising an
CC   isolated polynucleotide comprising a nucleotide sequence encoding a first
CC   polypeptide of at least 200 amino acids that has at least 90% identity
CC   based on the Clustal method of alignment when compared to a polypeptide
CC   comprising a sequence of 701, 541, 757, 345, 386, 330, 523, or 467 amino
CC   acids (EVEN SEQ ID NOS: 2-16), or an isolated polynucleotide comprising
CC   the complement of the nucleotide sequence, or comprising a polypeptide of
CC   at least 200 amino acids that has at least 90% identity based on the
CC   Clustal method of alignment when compared to a polypeptide of EVEN SEQ ID
CC   NOS: 2-16. Also described are: a chimeric gene comprising the isolated
CC   polynucleotide described above operably linked to suitable regulatory
CC   sequences; an isolated host cell comprising the chimeric gene of (1) or
CC   the isolated polynucleotide described above; a virus comprising the
CC   isolated polynucleotide described above; selecting an isolated
CC   polynucleotide that affects the level of expression of a polypeptide in a
CC   host or plant cell; obtaining a nucleic acid fragment encoding a
CC   polypeptide; and evaluating at least one compound for its ability to
CC   inhibit the activity of a polypeptide. The composition, methods,
CC   polynucleotides and polypeptides are useful in altering squalene
CC   metabolism in plants and seeds. This sequence encodes corn cycloartenol
CC   synthase, a member of the squalene metabolism pathway.
```

```
Db       1925 TTGATACTGTCACTTCAGAGTAAAAATGGAGGTATGACTGCGTGGGAGCCCGCAGGAGCT 1984

Qy        538 TyrAsnTrpLeuGluIleIleAsnProAlaGluThrPheGlyAspIleValIleAspTyr 557

Db       1985 TATAAATGGTTGGAACTACTCAATCCTACGGAATTTTTTGCTGACATAGTAGTTGAGCAC 2044

Qy        558 ProTyrValGluCysThrSerAlaAlaIleGlnAlaLeuAlaSerPheArgLysLeuTyr 577

Db       2045 GAATATCTTGAATGCACTGCATCAGCAATTCAGGTTTTAGTGTTGTTCAAAAAGCTTTAC 2104

Qy        578 ProGlyHisArgArgGluGluIleGlnHisCysIleAspLysAlaThrThrPheIleGlu 597

Db       2105 CCTGAGCATAGAAAGGAAGAGATAGAGAACTTCATTGCTAAAGCAGTAACATTCATTGAA 2164

Qy        598 LysIleGlnAlaSerAspGlySerTrpTyrGlySerTrpGlyValCysPheThrTyrGly 617

Db       2165 GATACACAATTAGAGAATGGTTCTTGGTATGGGAATTGGGCAGTTTGTTTCACTTACAGC 2224

Qy        618 AlaTrpPheGlyValLysGlyLeuIleAlaAlaGlyArgSerPheSerAsnCysSerSer 637

Db       2225 TCTTGGTTTGCACTTGGAGGTCTAGTTGCTGCTGGCAAGACTTACACAAATTGTGTTACT 2284

Qy        638 IleArgLysAlaCysGluPheLeuLeuSerLysGlnLeuProSerGlyGlyTrpGlyGlu 657

Db       2285 ATTCGTAAAGCTGTGAAATTTCTACTCAAAATACAAAATAAGGACGGTGGGTGGGGAGAG 2344

Qy        658 SerTyrLeuSerCysGlnAsnLysValTyrSerAsnLeuGluGlyAsnArgSerHisVal 677

Db       2345 AGTTATCTTTCTTGCCCAAGGAAGATGTACGTACCTCTTGAAGGAAGTCGATCAAATGTT 2404

Qy        678 ValAsnThrGlyTrpAlaMetLeuAlaLeuIleAspAlaGlyGlnAlaLysArgAspSer 697

Db       2405 GTACAAACATCATGGGCTCTAATGGCTCTAATTCATGCTGAGCAGGCTGAGAGAGATCCA 2464

Qy        698 GlnProLeuHisArgAlaAlaAlaTyrLeuIleAsnSerGlnLeuGluAspGlyAspPhe 717

Db       2465 ACTCCCCTTCATCATGCAGCAAAGTTACTCATTAATTCTCAGTTAGAAGATGGCGATTGG 2524

Qy        718 ProGlnGlnGluIleMetGlyValPheAsnLysAsnCysMetIleThrTyrAlaAlaTyr 737

Db       2525 CCCCAACAAGAAACTCTTGGAGTATACTTGAGAAATTGCTTGGTTCATTACTCATTCTAT 2584

Qy        738 ArgAsnIleProIleTrpAlaLeuGlyGluTyrGlnSerGlnValLeu 754

Db       2585 AGAAATATTTTTCCAATGTGGGCTTTGGCTGAATACCGCACAAATGTTTTA 2635
```

RESULT 13
ADE28074
ID ADE28074 standard; cDNA; 1882 BP.
XX
AC ADE28074;
XX
DT 29-JAN-2004 (first entry)
XX
DE Rice cycloartenol synthase cDNA.
XX
KW squalene metabolism; transgenic; plant; genetic; physical gene mapping;
KW breeding; rice; cycloartenol synthase; ss; gene.
XX
OS Oryza sativa.
XX

FH	Key	Location/Qualifiers
FT	CDS	3. .1628
FT		/*tag= a
FT		/partial
FT		/product= "Rice cycloartenol synthase protein"
FT		/note= "No start codon"

XX
PN US6630617-B1.
XX
PD 07-OCT-2003.
XX
PF 15-OCT-1999; 99US-00419679.
XX
PR 23-OCT-1998; 98US-0105405P.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Famodu OO, Kinney AJ, Rafalski JA;
XX
DR WPI; 2003-810554/76.
DR P-PSDB; ADE28075.
XX
PT New isolated polynucleotide encoding a squalene metabolic enzyme, useful
PT as probes in gene mapping, in producing transgenic plants or in plant
PT breeding to develop lines having desired phenotypes.
XX
PS Disclosure; SEQ ID NO 3; 31pp; English.
XX
CC The invention relates to a novel isolated polynucleotide encoding a
CC squalene metabolic enzyme. The polynucleotide of the invention may be
CC used to create transgenic plants in which the polypeptides are present at
CC higher or lower levels than normal or are found within cell types or
CC developmental stages in which they are not normally found. Furthermore,
CC the polynucleotide may be used as a probe for genetically and physically
CC mapping genes and as a marker for traits linked to those genes, as well
CC as in plant breeding in order to develop lines with desired phenotypes.
CC The current sequence is that of the rice cycloartenol synthase cDNA of
CC the invention.
XX
SQ Sequence 1882 BP; 565 A; 363 C; 441 G; 513 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	9.73e-230	Length:	1882
Score:	2433.00	Matches:	419
Percent Similarity:	90.43%	Conservative:	63
Best Local Similarity:	78.61%	Mismatches:	51
Query Match:	58.65%	Indels:	0
DB:	10	Gaps:	0

US-09-419-679-6 (1-757) x ADE28074 (1-1882)

```
Qy        222 LeuSerValLeuGlyValTyrGluTrpSerGlyAsnAsnProLeuProProGluIleTrp 241

Db         15 CTTTCGGTGCTTGGTGTATTTGACTGGTCTGGCAACAACCCAGTGCCACCAGAAATATGG 74

Qy        242 LeuLeuProTyrMetLeuProPheHisProGlyArgMetTrpCysHisCysArgMetVal 261

Db         75 TTGTTGCCATATTTCCTGCCGATTCATCCAGGGCGAATGTGGTGTCATTGCCGGATGGTT 134

Qy        262 TyrLeuProMetSerTyrLeuTyrGlyLysArgPheValGlyProIleSerProThrVal 281

Db        135 TATTTGCCTATGTGTTACATTTATGGAAAGAGGTTTGTGGGCCCAGTTACACCAATTATA 194

Qy        282 LeuSerLeuArgLysGluLeuTyrThrValProTyrHisAspIleAspTrpAspGlnAla 301

Db        195 TTGGAATTAAGAAAGGAACTCTACGAAGTACCCTACAATGAAGTTGATTGGGACAAGGCT 254

Qy        302 ArgAsnLeuCysAlaLysGluAspLeuTyrTyrProHisProLeuValGlnAspIleLeu 321

Db        255 CGCAATCTATGTGCTAAGGAAGATCTGTACTATCCACATCCATTCGTGCAGGATGTATTA 314

Qy        322 TrpAlaSerLeuHisLysPheLeuGluProIleLeuMetHisTrpProGlyLysArgLeu 341

Db        315 TGGGCCACTCTCCACAAATTTGTTGAACCAGCTATGTTGCGTTGGCCTGGGAACAAATTG 374

Qy        342 ArgGluLysAlaIleIleSerAlaLeuGluHisIleHisTyrGluAspGluAsnThrArg 361

Db        375 AGGGAGAAAGCTTTGGACACTGTCATGCAGCATATTCATTATGAAGATGAGAACACCCGA 434

Qy        362 TyrIleCysIleGlyProValAsnLysValLeuAsnMetLeuCysCysTrpValGluAsp 381

Db        435 TATATTTGCATTGGTCCAGTAAACAAGGTATTAAATATGCTTGCTTGCTGGATTGAAGAT 494

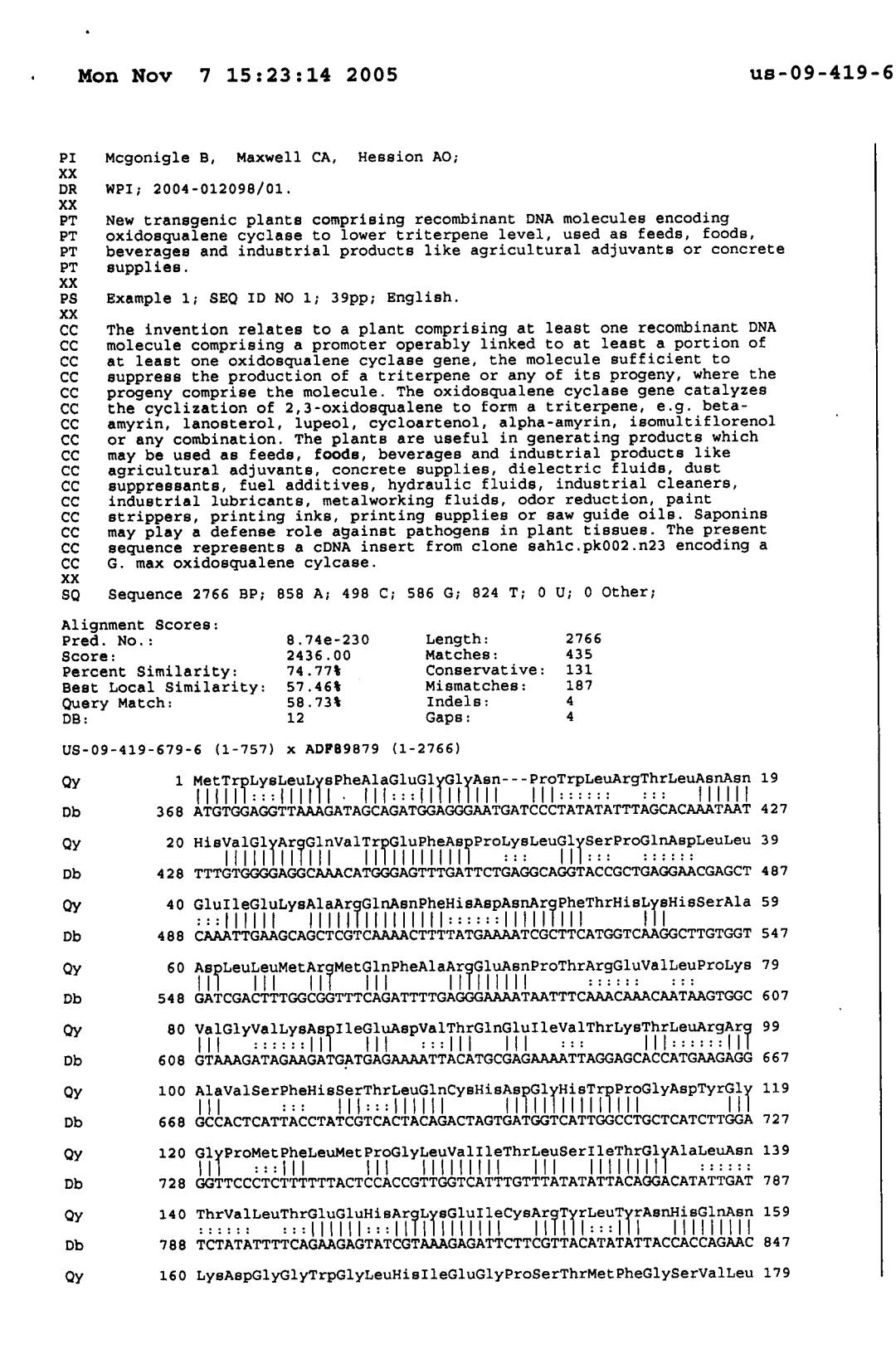
Qy        382 ProAsnSerGluAlaPheLysLeuHisLeuProArgIleTyrAspTyrLeuTrpIleAla 401
```

PI Mcgonigle B, Maxwell CA, Hession AO;
XX
DR WPI; 2004-012098/01.
XX
PT New transgenic plants comprising recombinant DNA molecules encoding
PT oxidosqualene cyclase to lower triterpene level, used as feeds, foods,
PT beverages and industrial products like agricultural adjuvants or concrete
PT supplies.
XX
PS Example 1; SEQ ID NO 1; 39pp; English.
XX
CC The invention relates to a plant comprising at least one recombinant DNA
CC molecule comprising a promoter operably linked to at least a portion of
CC at least one oxidosqualene cyclase gene, the molecule sufficient to
CC suppress the production of a triterpene or any of its progeny, where the
CC progeny comprise the molecule. The oxidosqualene cyclase gene catalyzes
CC the cyclization of 2,3-oxidosqualene to form a triterpene, e.g. beta-
CC amyrin, lanosterol, lupeol, cycloartenol, alpha-amyrin, isomultiflorenol
CC or any combination. The plants are useful in generating products which
CC may be used as feeds, foods, beverages and industrial products like
CC agricultural adjuvants, concrete supplies, dielectric fluids, dust
CC suppressants, fuel additives, hydraulic fluids, industrial cleaners,
CC industrial lubricants, metalworking fluids, odor reduction, paint
CC strippers, printing inks, printing supplies or saw guide oils. Saponins
CC may play a defense role against pathogens in plant tissues. The present
CC sequence represents a cDNA insert from clone sah1c.pk002.n23 encoding a
CC G. max oxidosqualene cylcase.
XX
SQ Sequence 2766 BP; 858 A; 498 C; 586 G; 824 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	8.74e-230	Length:	2766
Score:	2436.00	Matches:	435
Percent Similarity:	74.77%	Conservative:	131
Best Local Similarity:	57.46%	Mismatches:	187
Query Match:	58.73%	Indels:	4
DB:	12	Gaps:	4

US-09-419-679-6 (1-757) x ADF89879 (1-2766)



```
Qy       160 LysAspGlyGlyTrpGlyLeuHisIleGluGlyProSerThrMetPheGlySerValLeu 179
Db       848 AAAGATGGAGGTTGGGGACTACACATAGAAGGTCACAGTATCATGTTTTGCACTACACTC 907

Qy       180 SerTyrIleThrLeuArgLeuLeuGlyGluGlyProAsnAspGlyGlnGlyGlu---Met 198
Db       908 AATTATATATGCATGCGAATTCTTGGAGAAGGACCTAATGGAGGTCATAACAATGCTTGT 967

Qy       199 GluLysAlaArgAspTrpIleLeuGlyHisGlyGlyAlaThrTyrIleThrSerTrpGly 218
Db       968 GCTAAAGCAAGAAAGTGGATTCATGATCATGGTGGTGCAACACATATACCTTCATGGGGG 1027

Qy       219 LysMetTrpLeuSerValLeuGlyValTyrGluTrpSerGlyAsnAsnProLeuProPro 238
Db      1028 AAATTTTGGCTTTCGGTACTTGGTATAGTTGATTGGTGTGGAAGCAACCCAATGCCGCCT 1087

Qy       239 GluIleTrpLeuLeuProTyrMetLeuProPheHisProGlyArgMetTrpCysHisCys 258
Db      1088 GAATTTTGGATCCTTCCTTCTTTTCTCCCTATGCATCCGGGTAAAATGTGGTGTTATTGT 1147

Qy       259 ArgMetValTyrLeuProMetSerTyrLeuTyrGlyLysArgPheValGlyProIleSer 278
Db      1148 CGGTTGGTATACATGCCCATGTCTTATTTGTATGGGAAGAAATTTACGGGTCCAATCACA 1207

Qy       279 ProThrValLeuSerLeuArgLysGluLeuTyrThrValProTyrHisAspIleAspTrp 298
Db      1208 CCGTTAGTTGTAAATTTGAGAGAAGAACTTTTATTCAACCTTATGATGAAAATAGTTGG 1267

Qy       299 AspGlnAlaArgAsnLeuCysAlaLysGluAspLeuTyrTyrProHisProLeuValGln 318
Db      1268 AAGAAAGCACGTCATAAATGTGCAAATGAAGATCTTTACTATCCCCATCATTGGATACAA 1327

Qy       319 AspIleLeuTrpAlaSerLeuHisLysPheLeuGluProIleLeuMetHisTrpPro--- 337
Db      1328 GATCTATTATGGGATAGTTTGTATGTATTCACCGAGCCTCTTCTAAATTGTTGGCCTTTC 1387

Qy       338 GlyLysArgLeuArgGluLysAlaIleIleSerAlaLeuGluHisIleHisTyrGluAsp 357
Db      1388 AACAAGTTGGTTAGAGAAAAGGCACTTCAAGTAACAATGAAACATATTCATTATGAAGAC 1447

Qy       358 GluAsnThrArgTyrIleCysIleGlyProValAsnLysValLeuAsnMetLeuCysCys 377
Db      1448 GAAAATAGTCGGTATATTGCCATCGGGTGTGTGGAAAAGGTTCTATGTATGCTTGCTTGT 1507

Qy       378 TrpValGluAspProAsnSerGluAlaPheLysLeuHisLeuProArgIleTyrAspTyr 397
Db      1508 TGGGTTGAAGATCCAAATGGAGATGCTTTCAAGAAGCATCTTGCAAGGATCCCAGATTAT 1567

Qy       398 LeuTrpIleAlaGluAspGlyMetLysMetGlnGlyTyrAsnGlySerGlnLeuTrpAsp 417
Db      1568 TTATGGGTTTCTGAAGATGGAATGACCATGCAGGGTATT---GGTACTCAATCATGGGAT 1624

Qy       418 ThrAlaPheAlaValGlnAlaIleIleAlaSerAsnLeuIleGluGluPheGlyProThr 437
Db      1625 GTTGGTTTCATTGTTCAAGCTTTACTTGCTACTAACCTTATAGATGATTTTGGACCTACA 1684

Qy       438 IleArgLysAlaHisThrTyrIleLysAsnSerGlnValLeuGluAspCysProGlyAsp 457
Db      1685 ATTGCAAAAGCTCACGATTTCATCAAGAAATCTCAGGTAAGAGAAAATCCTTCGGGAGAT 1744

Qy       458 LeuAsnLysTrpTyrArgHisIleSerLysGlyAlaTrpProPheSerThrGlyAspHis 477
Db      1745 TTTAAGAGTATGTATCGTCACATTTGTAAAGGCTCATGGACCCTTGCCGATAGAGATCAT 1804

Qy       478 GlyTrpProIleSerAspCysThrAlaGluGlyLeuLysAlaValLeuLeuLeuSerLys 497
Db      1805 GCATGGCAAGTTTCTGATACCACTGCAGAATGTTTGAAGTGTTGTCTACTTTTATCAGTG 1864

Qy       498 IleAlaProGluIleValGlyGluProIleAspValLysArgLeuTyrAspSerValAsn 517
Db      1865 CTGCCACAAGATATTGTGGGAGAAAAAATGGAACTTGAAAAGTTACATGATTCAATCAAT 1924

Qy       518 ValIleLeuSerLeuGlnAsnGluAspGlyGlyPheAlaThrTyrGluLeuLysArgSer 537
Db      1925 TTGATACTGTCACTTCAGAGTAAAAATGGAGGTATGACTGCGTGGGAGCCCGCAGGAGCT 1984

Qy       538 TyrAsnTrpLeuGluIleIleAsnProAlaGluThrPheGlyAspIleValIleAspTyr 557
Db      1985 TATAAATGGTTGGAACTACTCAATCCTACGGAATTTTTTGCTGACATAGTAGTTGAGCAC 2044

Qy       558 ProTyrValGluCysThrSerAlaAlaIleGlnAlaLeuAlaSerPheArgLysLeuTyr 577
Db      2045 GAATATCTTGAATGCACTGCATCAGCAATTCAGGTTTTAGTGTTGTTCAAAAAGCTTTAC 2104

Qy       578 ProGlyHisArgArgGluGluIleGlnHisCysIleAspLysAlaThrThrPheIleGlu 597
Db      2105 CCTGAGCATAGAAAGGAAGAGATAGAGAACTTCATTGCTAAAGCAGTAACATTCATTGAA 2164

Qy       598 LysIleGlnAlaSerAspGlySerTrpTyrGlySerTrpGlyValCysPheThrTyrGly 617
Db      2165 GATACACAATTAGAGAATGGTTCTTGGTATGGGAATTGGGCAGTTTGTTTCACTTACAGC 2224

Qy       618 AlaTrpPheGlyValLysGlyLeuIleAlaAlaGlyArgSerPheSerAsnCysSerSer 637
Db      2225 TCTTGGTTTGCACTTGGAGGTCTAGTTGCTGCTGGCAAGACTTACACAAATTGTGTTACT 2284

Qy       638 IleArgLysAlaCysGluPheLeuLeuSerLysGlnLeuProSerGlyGlyTrpGlyGlu 657
Db      2285 ATTCGTAAAGCTGTGAAATTTCTACTCAAAATACAAAATAAGGACGGTGGGTGGGGAGAG 2344

Qy       658 SerTyrLeuSerCysGlnAsnLysValTyrSerAsnLeuGluGlyAsnArgSerHisVal 677
Db      2345 AGTTATCTTTCTTGCCCAAGGAAGATGTACGTACCTCTTGAAGGAAGTCGATCAAATGTT 2404

Qy       678 ValAsnThrGlyTrpAlaMetLeuAlaLeuIleAspAlaGlyGlnAlaLysArgAspSer 697
Db      2405 GTACAAACATCATGGGCTCTAATGGCTCTAATTCATGCTGAGCAGGCTGAGAGAGATCCA 2464

Qy       698 GlnProLeuHisArgAlaAlaAlaTyrLeuIleAsnSerGlnLeuGluAspGlyAspPhe 717
Db      2465 ACTCCCCTTCATCATGCAGCAAAGTTACTCATTAATTCTCAGTTAGAAGATGGCGATTGG 2524

Qy       718 ProGlnGlnGluIleMetGlyValPheAsnLysAsnCysMetIleThrTyrAlaAlaTyr 737
Db      2525 CCCCAACAAGAAACTCTTGGAGTATACTTGAGAAATTGCTTGGTTCATTACTCATTCTAT 2584

Qy       738 ArgAsnIlePheProIleTrpAlaLeuGlyGluTyrGlnSerGlnValLeu 754
Db      2585 AGAAATATTTTTCCAATGTGGGCTTTGGCTGAATACCGCACAAATGTTTTA 2635
```

RESULT 12
ADF89879
ID ADF89879 standard; cDNA; 2766 BP.
XX
AC ADF89879;
XX
DT 26-FEB-2004 (first entry)
XX
DE G. max oxidosqualene cylcase cDNA fragment (clone sah1c.pk002.n23).
XX
KW Oxidosqualene cyclase; triterpene; plant protectant; industrial product;
KW plant; soybean; gene; ss.
XX
OS Glycine max.
XX
PN WO2003095615-A2.
XX
PD 20-NOV-2003.
XX
PF 05-MAY-2003; 2003WO-US014218.
XX
PR 09-MAY-2002; 2002US-0379361P.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX

```
Db       1956 GAATATGTTGAGTGCACTGGATCTGCAATTCAAGCTTTAGTTTTGTTCCAGAAGCTATAT 2015

Qy        578 ProGlyHisArgArgGluGluIleGlnHisCysIleAspLysAlaThrThrPheIleGlu 597

Db       2016 CCAGAGCATAGGAAGACAGAGATTAAGAATTTCATTGTCAATGCAGTTCAATTCCTTGAA 2075

Qy        598 LysIleGlnAlaSerAspGlySerTrpTyrGlySerTrpGlyValCysPheThrTyrGly 617

Db       2076 GATACACAAACAACCAATGGTTCATGGTATGGATGTTGGGGAGTTTGCTACACTTATGGC 2135

Qy        618 AlaTrpPheGlyValLysGlyLeuIleAlaAlaGlyArgSerPheSerAsnCysSerSer 637

Db       2136 TCTTGGTTTGCACTTGGTGGTCTAGCAGCTGCCGGTAAGACTTACACTAATTGTAATGCC 2195

Qy        638 IleArgLysAlaCysGluPheLeuLeuSerLysGlnLeuProSerGlyGlyTrpGlyGlu 657

Db       2196 ATTCGCAAGGCTGTGAAATTTCTACTTACAACACAGAGAGAGGATGGTGGGTGGGGAGAG 2255

Qy        658 SerTyrLeuSerCysGlnAsnLysValTyrSerAsnLeuGluGlyAsnArgSerHisVal 677

Db       2256 AGTTATCTTTCAAGCCCAAAAAAGATATACATACCTCTTGAAGGAAGTCGATCAAACGTT 2315

Qy        678 ValAsnThrGlyTrpAlaMetLeuAlaLeuIleAspAlaGlyGlnAlaLysArgAspSer 697

Db       2316 GTACAAACAGCATGGGCTCTTATGGGTCTAATTTATGCTGGACAGTCAGAGAGAGACCTT 2375

Qy        698 GlnProLeuHisArgAlaAlaAlaTyrLeuIleAsnSerGlnLeuGluAspGlyAspPhe 717

Db       2376 ACTCCTCTTCATCGAGCTGCAAAGTTGCTCATTAATTCCCAGTTGGAAGAAGGTGATTGG 2435

Qy        718 ProGlnGlnGluIleMetGlyValPheAsnLysAsnCysMetIleThrTyrAlaAlaTyr 737

Db       2436 CCCCAACAGGAAATCACTGGAGTATTCTTGAAAAACTGCATGATGCATTACCCAATGTAT 2495

Qy        738 ArgAsnIleProIleTrpAlaLeuGlyGluTyrGlnSerGlnVal 753

Db       2496 AGAAATATTTTTCCAATGTGGGCTCTAGCTGAATATCGTAGGCGAGTT 2543
```

RESULT 11
AAD16895
ID AAD16895 standard; cDNA; 2766 BP.
XX
AC AAD16895;
XX
DT 29-NOV-2001 (first entry)
XX
DE Soybean oxidosqualene cyclase protein cDNA from clone sah1c.pk002.n23.
XX
KW Soybean; oxidosqualene cyclase; cytostatic; antilipaemic; colon cancer;
KW transgenic plant; insect tolerance; soy flour; bean; nut; tofu; natto;
KW miso; tempeh; ss.
XX
OS Glycine max.
XX
FH Key Location/Qualifiers
FT CDS 368. .2656
FT /*tag= a
FT /product= "Soybean oxidosqualene cyclase protein"
XX
PN WO200166773-A2.
XX
PD 13-SEP-2001.
XX
PF 09-MAR-2001; 2001WO-US007611.
XX
PR 09-MAR-2000; 2000US-0188054P.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Bryan GT, Mcgonigle B, Maxwell CA, Potter SM, Hwang D;
XX
DR WPI; 2001-565594/63.
DR P-PSDB; AAE09796.
XX
PT New nucleic acid fragment encoding oxidosqualene cyclase, useful as DNA
PT hybridization probe, polymerase chain reaction primer and for creating
PT transgenic plants with cells having higher or lower of oxidosqualene
PT cyclase.
XX
PS Claim 2; Page 47-48; 59pp; English.
XX
CC The invention relates to oxidosqualene cyclase proteins and their
CC corresponding cDNA molecules. Oxidosqualene cyclase genes are used as
CC hybridisation probes to screen libraries from any desired plant.
CC Oxidosqualene cyclase DNA are used to create transgenic plants with cells
CC having higher or lower level of oxidosqualene cyclase. These changes in
CC the plant seed are useful to improve the seed nutritional value, and in
CC the plant leaf may aid in insect tolerance. Oxidosqualene cyclase are
CC produced in heterologous host cells and are used to prepare antibodies.
CC The antibodies are useful for detecting oxidosqualene cyclase in situ in
CC cells or in vitro in cell extracts. The invention is useful for altering
CC level of expression of oxidosqualene cyclase (beta-amyrin synthase)
CC protein in a plant cell. Soybean grain produced from a transgenic plant
CC is useful for preparing food products such as soy protein product,
CC soybean meal, soy flour, soy protein concentrate, soy milk, a dietary
CC supplement, beans, nuts, tofu, natto, miso and tempeh. The formed soy
CC protein product having increased saponin levels is useful for lowering
CC cholesterol in a person suspected of having increased cholesterol, and
CC for decreasing risk of colon cancer in a person suspected of having risk
CC of cancer. The present cDNA sequence encodes soybean oxidosqualene
CC cyclase protein
XX
SQ Sequence 2766 BP; 858 A; 498 C; 586 G; 824 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	8.74e-230	Length:	2766
Score:	2436.00	Matches:	435
Percent Similarity:	74.77%	Conservative:	131
Best Local Similarity:	57.46%	Mismatches:	187
Query Match:	58.73%	Indels:	4
DB:	4	Gaps:	4

US-09-419-679-6 (1-757) x AAD16895 (1-2766)

```
Qy          1 MetTrpLysLeuLysPheAlaGluGlyGlyAsn---ProTrpLeuArgThrLeuAsnAsn 19

Db        368 ATGTGGAGGTTAAAGATAGCAGATGGAGGGAATGATCCCTATATATTTAGCACAAATAAT 427

Qy         20 HisValGlyArgGlnValTrpGluPheAspProLysLeuGlySerProGlnAspLeuLeu 39

Db        428 TTTGTGGGGAGGCAAACATGGGAGTTTGATTCTGAGGCAGGTACCGCTGAGGAACGAGCT 487

Qy         40 GluIleGluLysAlaArgGlnAsnPheHisAspAsnArgPheThrHisLysHisSerAla 59

Db        488 CAAATTGAAGCAGCTCGTCAAAACTTTTATGAAAATCGCTTCATGGTCAAGGCTTGTGGT 547

Qy         60 AspLeuLeuMetArgMetGlnPheAlaArgGluAsnProThrArgGluValLeuProLys 79

Db        548 GATCGACTTTGGCGGTTTCAGATTTTGAGGGAAAATAATTTCAAACAAACAATAAGTGGC 607

Qy         80 ValGlyValLysAspIleGluAspValThrGlnGluIleValThrLysThrLeuArgArg 99

Db        608 GTAAAGATAGAAGATGATGAGAAAATTACATGCGAGAAAATTAGGAGCACCATGAAGAGG 667

Qy        100 AlaValSerPheHisSerThrLeuGlnCysHisAspGlyHisTrpProGlyAspTyrGly 119

Db        668 GCCACTCATTACCTATCGTCACTACAGACTAGTGATGGTCATTGGCCTGCTCATCTTGGA 727

Qy        120 GlyProMetPheLeuMetProGlyLeuValIleThrLeuSerIleThrGlyAlaLeuAsn 139

Db        728 GGTTCCCTCTTTTTTACTCCACCGTTGGTCATTTGTTTATATATTACAGGACATATTGAT 787

Qy        140 ThrValLeuThrGluGluHisArgLysGluIleCysArgTyrLeuTyrAsnHisGlnAsn 159

Db        788 TCTATATTTTCAGAAGAGTATCGTAAAGAGATTCTTCGTTACATATATTACCACCAGAAC 847
```

CC The invention relates to oxidosqualene cyclase proteins and their
CC corresponding cDNA molecules. Oxidosqualene cyclase genes are used as
CC hybridisation probes to screen libraries from any desired plant.
CC Oxidosqualene cyclase DNA are used to create transgenic plants with cells
CC having higher or lower level of oxidosqualene cyclase. These changes in
CC the plant seed are useful to improve the seed nutritional value, and in
CC the plant leaf may aid in insect tolerance. Oxidosqualene cyclase are
CC produced in heterologous host cells and are used to prepare antibodies.
CC The antibodies are useful for detecting oxidosqualene cyclase in situ in
CC cells or in vitro in cell extracts. The invention is useful for altering
CC level of expression of oxidosqualene cyclase (beta-amyrin synthase)
CC protein in a plant cell. Soybean grain produced from a transgenic plant
CC is useful for preparing food products such as soy protein product,
CC soybean meal, soy flour, soy protein concentrate, soy milk, a dietary
CC supplement, beans, nuts, tofu, natto, miso and tempeh. The formed soy
CC protein product having increased saponin levels is useful for lowering
CC cholesterol in a person suspected of having increased cholesterol, and
CC for decreasing risk of colon cancer in a person suspected of having risk
CC of cancer. The present cDNA sequence encodes soybean oxidosqualene
CC cyclase protein
XX
SQ Sequence 2731 BP; 877 A; 458 C; 601 G; 795 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.27e-231	Length:	2731
Score:	2454.50	Matches:	437
Percent Similarity:	74.74%	Conservative:	128
Best Local Similarity:	57.80%	Mismatches:	186
Query Match:	59.17%	Indels:	5
DB:	4	Gaps:	5

US-09-419-679-6 (1-757) x AAD16893 (1-2731)

```
Qy          1 MetTrpLysLeuLysPheAlaGluGlyGly---AsnProTrpLeuArgThrLeuAsnAsn 19
              ||||||:::|||||   |||:::||||||   :::|||:::::   :::   ||||||
Db        282 ATGTGGAGGCTGAAGATAGCAGATGGTGGAAAAGACCCATACATTTTCAGCACAAACAAT 341

Qy         20 HisValGlyArgGlnValTrpGluPheAspProLysLeuGlySerProGlnAspLeuLeu 39
                ||||||||||||   ||||||||||||||:::   |||:::|||::::::
Db        342 TTTGTAGGAAGACAGACATGGGAGTTTGATCCTGAGGCAGGAACTCCAGAGGAACGAGCT 401

Qy         40 GluIleGluLysAlaArgGlnAsnPheHisAspAsnArgPheThrHisLysHisSerAla 59
              ::::::|||   ||||||||||||||:::::|||||||||      |||
Db        402 CAAGTTGAAGCAGCTCGTCAAAATTTTTATAACAATCGCTTCAAGGTCAAGGCATGTGGT 461

Qy         60 AspLeuLeuMetArgMetGlnPheAlaArgGluAsnProThrArgGluValLeuProLys 79
              |||||||||   |||   |||   ||||||          ::::::   :::|||
Db        462 GATCTCCTTTGGCGTTTTCAGATTCTGAGAGAAAAAAACTTCAAACAATCAATACCTAGT 521

Qy         80 ValGlyValLysAspIleGluAspValThrGlnGluIleValThrLysThrLeuArgArg 99
              |||   :::::::|||   |||:::::|||   |||   |||       ||||||||||||
Db        522 GTGAAGATAGAAGATGGAGAAGAAATAACATACGAAAAAGTCATAAGCACGTTGAGAAGA 581

Qy        100 AlaValSerPheHisSerThrLeuGlnCysHisAspGlyHisTrpProGlyAspTyrGly 119
              |||              |||   ||||||      ||||||||||||||
Db        582 GCCGCACACCACCTATCAGCATTGCAGACCAGTGATGGCCATTGGCCTGCACAAATTGCA 641

Qy        120 GlyProMetPheLeuMetProGlyLeuValIleThrLeuSerIleThrGlyAlaLeuAsn 139
              ||||||:::|||   :::|||   ||||||      :::   |||||||||   |||:::
Db        642 GGTCCTCTGTTTTTTCTGCCTCCTTTGGTTTTTGTATGTACATTACAGGGCATCTTGAT 701

Qy        140 ThrValLeuThrGluGluHisArgLysGluIleCysArgTyrLeuTyrAsnHisGlnAsn 159
                 |||      ||||||:::||||||||||||   ||||||::|||   |||||||||
Db        702 TTGGTATTCCCAGAAGAGTATCGAAAAGAGATTCTCCGTTACATATACTATCACCAGAAT 761

Qy        160 LysAspGlyGlyTrpGlyLeuHisIleGluGlyProSerThrMetPheGlySerValLeu 179
              :::||||||||||||||||||||||||||||||   ||||||||||||||:::   |||
Db        762 GAAGATGGAGGATGGGGACTACACATAGAGGGTCATAGCACCATGTTTGGTACTACACTA 821

Qy        180 SerTyrIleThrLeuArgLeuLeuGlyGluGlyProAsnAspGly---GlnGlyGluMet 198
              :::||||||   :::|||:::||||||||||||||||||   |||   :::
Db        822 AACTATATATGTATGCGAATTCTTGGAGAAGGGCCTAATGGAGGTCATGAAAATGCATGT 881

Qy        199 GluLysAlaArgAspTrpIleLeuGlyHisGlyGlyAlaThrTyrIleThrSerTrpGly 218
                 :::   :::   ||||||      |||||||||   |||:::|||   |||||||||
Db        882 GCTAGAGGAAAAAAGTGGATTCATGATCATGGTGGTGTAACACACATCCCTTCATGGGGG 941

Qy        219 LysMetTrpLeuSerValLeuGlyValTyrGluTrpSerGlyAsnAsnProLeuProPro 238
              |||   |||||||||:::|||||||||:::::|||   |||:::||||||:::||||||
Db        942 AAAACTTGGCTTTCGATACTTGGTGTATTTGATTGGTGTGGAAGCAACCCAATGCCCCCA 1001

Qy        239 GluIleTrpLeuLeuProTyrMetLeuProPheHisProGlyArgMetTrpCysHisCys 258
              |||   |||::::::|||      ||||||   ||||||   :::|||||||||:::|||
Db       1002 GAGTTTTGGATTATTCCTTCGTTTCTTCCTATGCATCCAGCTAAAATGTGGTGTTATTGT 1061

Qy        259 ArgMetValTyrLeuProMetSerTyrLeuTyrGlyLysArgPheValGlyProIleSer 278
              |||:::||||||:::|||||||||||:::   ||||||   ||||||||||||||:::
Db       1062 CGATTGGTATATATGCCTATGTCTTATATG---GGAAAAGGGTTTGTGGGTCCAATCACA 1118

Qy        279 ProThrValLeuSerLeuArgLysGluLeuTyrThrValProTyrHisAspIleAspTrp 298
              |||   :::|||   ||||||:::||||||:::|||   ||||||      :::::::|||
Db       1119 CCACTCATCTTACAATTGAGAGAAGAGCTCTTTACTCAGCCTTATGAAAAAGTTAATTGG 1178

Qy        299 AspGlnAlaArgAsnLeuCysAlaLysGluAspLeuTyrTyrProHisProLeuValGln 318
                 :::   |||:::   |||||||||||||||||||||||||||||   |||:::|||
Db       1179 AAAAAAGTGCGTCATCAATGTGCAAAGGAAGATCTTTATTATCCTCATTCTTTGATACAA 1238

Qy        319 AspIleLeuTrpAlaSerLeuHisLysPheLeuGluProIleLeuMetHisTrpPro--- 337
              |||:::::|||   |||||:::   |||   ||||||:::|||      ||||||
Db       1239 GACCTAGTATGGGACAGTCTATACATGTTTACTGAGCCACTACTTACTTGTTGGCCTTTC 1298

Qy        338 GlyLysArgLeuArgGluLysAlaIleIleSerAlaLeuGluHisIleHisTyrGluAsp 357
                 |||   :::|||||||||||:::         :::   ||||||||||||||||||
Db       1299 AACAAACTAATTAGAGAAAAGGCCCTTCAAGTAACAATGAACCATATTCATTACGAAGAT 1358

Qy        358 GluAsnThrArgTyrIleCysIleGlyProValAsnLysValLeuAsnMetLeuCysCys 377
              ||||||:::|||||||||   ||||||   |||   |||||||||   ||||||   |||
Db       1359 GAGAATAGTCGATACATAACTATTGGGTGTGTGGAAAAGGTTCTATGTATGCTTGCTTGT 1418

Qy        378 TrpValGluAspProAsnSerGluAlaPheLysLeuHisLeuProArgIleTyrAspTyr 397
              ||||||||||||||      :::|||||||||   ||||||   |||:::   ||||||
Db       1419 TGGGTTGAAGATCCACAGGGAGATGCCTTTAAGAAGCATCTTGCAAGGGTCTCAGATTAC 1478

Qy        398 LeuTrpIleAlaGluAspGlyMetLysMetGlnGlyTyrAsnGlySerGlnLeuTrpAsp 417
              ||||||::::::|||||||||||||   ||||||   :::   |||||||||   ||||||
Db       1479 TTATGGGTTTCTGAAGATGGAATGACCATGCAAAGTTTT---GGTAGTCAAGAATGGGAT 1535

Qy        418 ThrAlaPheAlaValGlnAlaIleIleAlaSerAsnLeuIleGluGluPheGlyProThr 437
                 ||||||||||||||:::::|||:::   ||||||:::|||   |||   :::
Db       1536 GCTGGTTTTGCAGTTCAAGCTTTGCTTGCAACTAAGCTAATTGACGAAATTGGCCATTCA 1595

Qy        438 IleArgLysAlaHisThrTyrIleLysAsnSerGlnValLeuGluAspCysProGlyAsp 457
              :::   |||   |||   :::||||||   |||||||||   ::::::      ||||||
Db       1596 CTTGCAAAAGGGCATGATTTCATCAAGAAGTCTCAGGTGAGAGACAACCCTTCAGGAGAT 1655

Qy        458 LeuAsnLysTrpTyrArgHisIleSerLysGlyAlaTrpProPheSerThrGlyAspHis 477
                    ||||||||||||:::||||||:::|||   ||||||      ||||||
Db       1656 TTTAAGAGTATGTATCGTCATATTACTAAAGGGTCTTGGACATTCTCAGATCAAGACCAT 1715

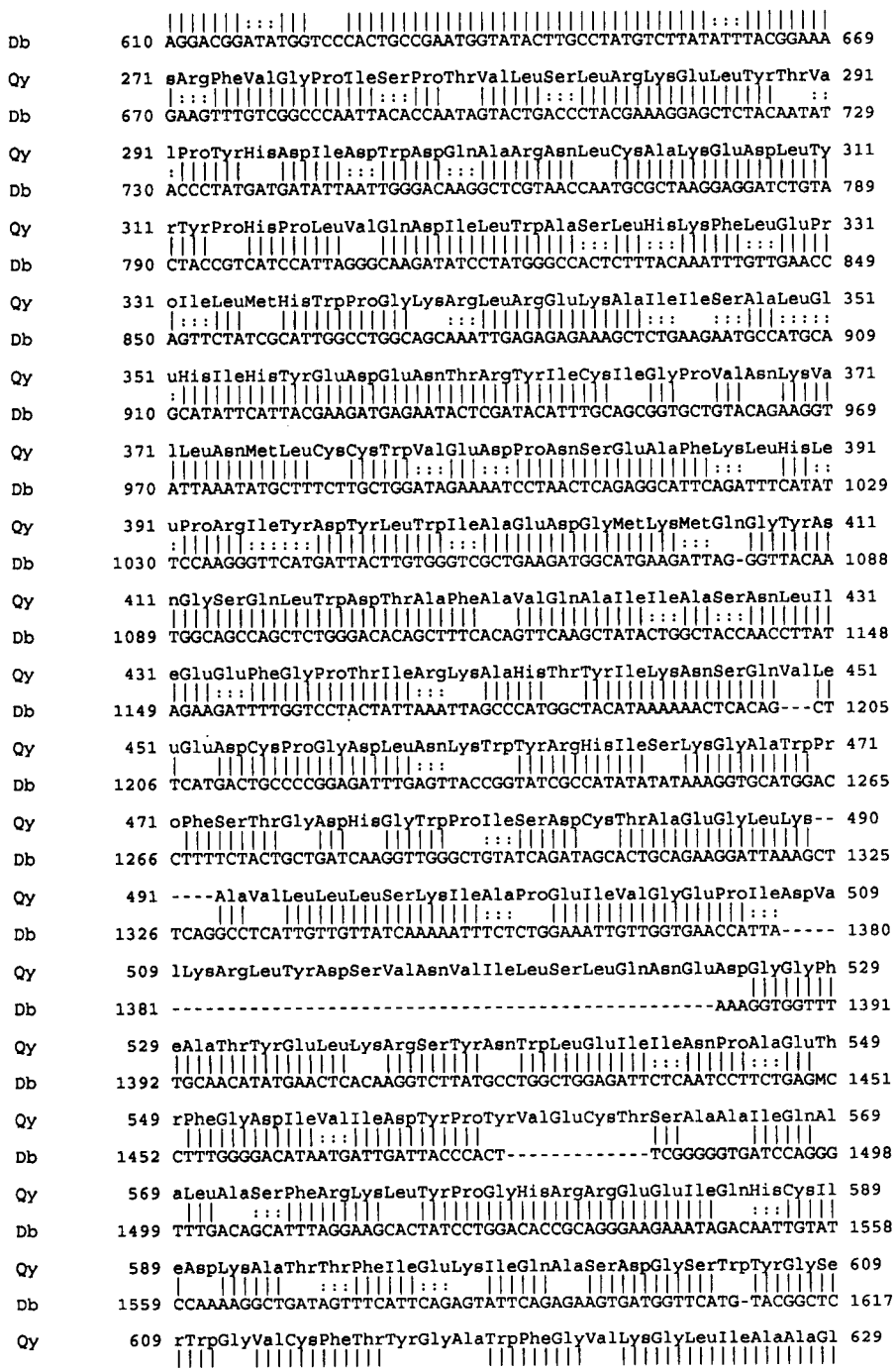
Qy        478 GlyTrpProIleSerAspCysThrAlaGluGlyLeuLysAlaValLeuLeuLeuSerLys 497
              ||||||   :::||||||||||||||||||||||||||||||      ||||||||||||||
Db       1716 GGATGGCAAGTTTCTGATTGCACTGCAGAAGGTTAAAGTGTTGTCTTCTTCTATCAAAG 1775

Qy        498 IleAlaProGluIleValGlyGluProIleAspValLysArgLeuTyrAspSerValAsn 517
              ::::::||||||||||||||||||   :::   :::|||   ||||||||||||||||
Db       1776 TTGTCACCAGAGATTGTGGGAGAAAAAGTGAAACCTGAAAGATTTTATGATTCAGTCAAT 1835

Qy        518 ValIleLeuSerLeuGlnAsnGluAspGlyGlyPheAlaThrTyrGluLeuLysArgSer 537
              ::::::||||||||||||::::::   ||||||   |||   :::|||          :::
Db       1836 ATCTTGTTGTCACTTCAGAGTAAAAAAGGTGGTATAGCAGCATGGGAACCAATAGGAGCT 1895

Qy        538 TyrAsnTrpLeuGluIleIleAsnProAlaGluThrPheGlyAspIleValIleAspTyr 557
                    |||||||||::::::||||||   |||   |||   ||||||||||||::::::
Db       1896 CAAGAATGGTTGGAATTACTCAATCCCACTGAATTTTTTGAGGACATTGTAATTGAGCAT 1955

Qy        558 ProTyrValGluCysThrSerAlaAlaIleGlnAlaLeuAlaSerPheArgLysLeuTyr 577
```



RESULT 10
AAD16893
ID AAD16893 standard; cDNA; 2731 BP.
XX
AC AAD16893;
XX
DT 29-NOV-2001 (first entry)
XX
DE Soybean oxidosqualene cyclase protein cDNA from clone sdp3c.pk020.o10.
XX
KW Soybean; oxidosqualene cyclase; cytostatic; antilipaemic; colon cancer;
KW transgenic plant; insect tolerance; soy flour; bean; nut; tofu; natto;
KW miso; tempeh; ss.
XX
OS Glycine max.
XX
FH Key Location/Qualifiers
FT CDS 282. .2567
FT /*tag= a
FT /product= "Soybean oxidosqualene cyclase protein"
XX
PN WO200166773-A2.
XX
PD 13-SEP-2001.
XX
PF 09-MAR-2001; 2001WO-US007611.
XX
PR 09-MAR-2000; 2000US-0188054P.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Bryan GT, Mcgonigle B, Maxwell CA, Potter SM, Hwang D;
XX
DR WPI; 2001-565594/63.
DR P-PSDB; AAE09794.
XX
PT New nucleic acid fragment encoding oxidosqualene cyclase, useful as DNA
PT hybridization probe, polymerase chain reaction primer and for creating
PT transgenic plants with cells having higher or lower of oxidosqualene
PT cyclase.
XX
PS Claim 2; Page 41-42; 59pp; English.
XX

XX
DT 22-APR-2004 (first entry)
XX
DE Rice isoprenoid biosynthesis-associated cDNA #58.
XX
KW Rice; isoprenoid biosynthesis; ss; gene; plant; isopentenyl diphosphate;
KW IPP; dimethylallyl alcohol; DMAPP; short-chain plastid prenyltransferase;
KW gibberellin; carotenoid; abscisic acid; tocopherol; plastoquinone;
KW phylloquinone; mevalonate pathway; phytosterol; brassinosteroid;
KW ubiquinone; monterpene; sesquiterpene; protein prenylation; chlorophyll;
KW haeme; yield.
XX
OS Oryza sativa.
XX
PN US2004010815-A1.
XX
PD 15-JAN-2004.
XX
PF 26-SEP-2002; 2002US-00259194.
XX
PR 26-SEP-2001; 2001US-0325277P.
PR 04-APR-2002; 2002US-0370620P.
PR 04-APR-2002; 2002US-0370743P.
XX
PA (LANG/) LANGE B M.
PA (GHAS/) GHASSEMIAN M.
PA (BRIG/) BRIGGS S P.
PA (COOP/) COOPER B.
PA (GLAZ/) GLAZEBROOK J.
PA (GOFF/) GOFF S A.
PA (KATA/) KATAGIRI F.
PA (KREP/) KREPS J.
PA (MOUG/) MOUGHAMER T.
PA (PROV/) PROVART N.
PA (RICK/) RICKE D.
PA (ZHUT/) ZHU T.
XX
PI Lange BM, Ghassemian M, Briggs SP, Cooper B, Glazebrook J;
PI Goff SA, Katagiri F, Kreps J, Moughamer T, Provart N, Ricke D;
PI Zhu T;
XX
DR WPI; 2004-090562/09.
DR P-PSDB; ADI45185.
XX
PT New isolated polynucleotides and polypeptides associated with isoprenoid
PT synthesis in plants, useful for producing transgenic plants, for targeted
PT gene disruption, as well as markers or probes.
XX
PS Claim 1; SEQ ID NO 115; 117pp; English.
XX
CC The invention relates to a polynucleotide (or its complement, protein
CC encoding fragment or reverse complement), comprising a nucleotide
CC sequence encoding a polypeptide comprising an amino acid sequence
CC involved in or associated with the biosynthesis of isoprenoids in a rice
CC plant. Also included are an isolated polypeptide involved in or
CC associated with the biosynthesis of isoprenoids in a plant, an expression
CC cassette comprising the polynucleotide, a host cell comprising the
CC expression cassette, and a transgenic plant comprising the expression
CC cassette. The polypeptides and polynucleotides include those associated
CC with the biosynthesis of isopentenyl diphosphate (IPP) and dimethylallyl
CC alcohol (DMAPP), the biosynthesis of short-chain plastid
CC prenyltransferases, the biosynthesis of gibberellins, the biosynthesis of
CC carotenoids and/or abscisic acids, the biosynthesis of tocopherols,
CC plastoquinone and/or phylloquinone biosynthesis, the mevalonate pathway,
CC phytosterol and brassinosteroid metabolism, biosynthesis of ubiquinone,
CC biosynthesis of monterpenes and sesquiterpenes, protein prenylation, and
CC biosynthesis of chlorophyll or haeme. Also disclosed are banana, wheat
CC and corn homologues of some of the rice polynucleotides. The
CC polynucleotides are useful for producing transgenic plants, where the
CC genome is augmented by a nucleic acid molecule of the invention, or in
CC which the corresponding gene has been disrupted, e.g. to result in a
CC loss, a decrease or an alteration in the function of the product encoded
CC by the gene. The plants may also have increased yields and/or produce a
CC better quality product than the corresponding wild-type plant. The
CC nucleic acid molecules are useful for targeted gene disruption, as well
CC as markers and probes. Note: The sequence data for this patent did not
CC form part of the printed specification, but was obtained in electronic
CC format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=20040010815. The present sequence
CC is a Rice isoprenoid biosynthesis- associated cDNA of the invention.
XX
SQ Sequence 2093 BP; 565 A; 446 C; 529 G; 552 T; 0 U; 1 Other;

Alignment Scores:

Pred. No.:	6.42e-238	Length:	2093
Score:	2516.50	Matches:	492
Percent Similarity:	74.09%	Conservative:	80
Best Local Similarity:	63.73%	Mismatches:	101
Query Match:	60.67%	Indels:	102
DB:	12	Gaps:	14

US-09-419-679-6 (1-757) x ADI45184 (1-2093)

```
Qy          1 MetTrpLysLeuLysPheAlaGlu-----------GlyGlyAsnProTrpLeuArgThr 16
              ||||||:::|||:::  ||||||           ||||||     ||||||  :::
Db          1 ATGTGGCGGCTAAGAGTCGCGGAGGAGGCCGGCGCCGGCGGCAAGGGGTGGCTGAGCTCC 60

Qy         17 LeuAsnAsnHisValGlyArgGlnValTrpGluPheAspProLysLeuGlySerProGln 36
                 |||  ||||||||||||||                  |||||||||
Db         61 ACCAACGCGCACGTCGGCCGGCAG-----------------AAGCTGGGC--------- 93

Qy         37 AspLeuLeuGluIleGluLysAlaArgGlnAsnPheHisAspAsnArgPheThrHisLys 56

Db         93 ------------------------------------------------------------ 93

Qy         57 HisSerAlaAspLeuLeuMetArgMetGlnPheAlaArgGluAsnProThrArgGluVal 76

Db         93 ------------------------------------------------------------ 93

Qy         77 LeuProLysValGlyValLysAspIleGluAspValThrGlnGluIleValThrLysThr 96
                          :::  ||||||||||||:::|||  |||     :::
Db         94 --------------------GAGGATGAAGATGTCACAGAAGAAGCGGTGCTGACCTCT 132

Qy         97 LeuArgArgAlaValSerPheHisSerThrLeuGlnCysHisAspGlyHisTrpProGly 116
              |||:::||||||:::      :::||||||||||||  ||||||||||||||||||||
Db        133 TTGAAAAGGGCGATTCGGCGTTATTCTACTCTTCAGGCACATGATGGACACTGGCCTGGG 192

Qy        117 AspTyrGlyGlyProMetPheLeuMetProGlyLeuValIleThrLeuSerIleThrGly 136
              ||||||  ||||||||||||||:::||||||||:::::   |||  :::|||||||
Db        193 GATTATGCTGGTCCCATGTTCCTTCTTCCAGGCTTGATAGTA---CTGCATGTGACTGGT 249

Qy        137 AlaLeuAsnThrValLeuThrGluGluHisArgLysGluIleCysArgTyrLeuTyrAsn 156
              ||||||||||||||||:::   ||||||:::|||||||||  ||||||||||||||
Db        250 GCACTAAATACTGTGCTATCAACAGAACATCAAAAGGAGATTCGCCGGTATCTCTATAAT 309

Qy        157 His---GlnAsnLysAspGlyGlyTrpGlyLeuHisIleGluGlyProSerThrMetPhe 175
              |||   ||||||:::||||||||||||||||||||||||||||  ||||||||||||
Db        310 CATCAGCAGAATGAAGATGGAGGCTGGGGGCTGCACATTGAGGGCACAAGCACCATGTTC 369

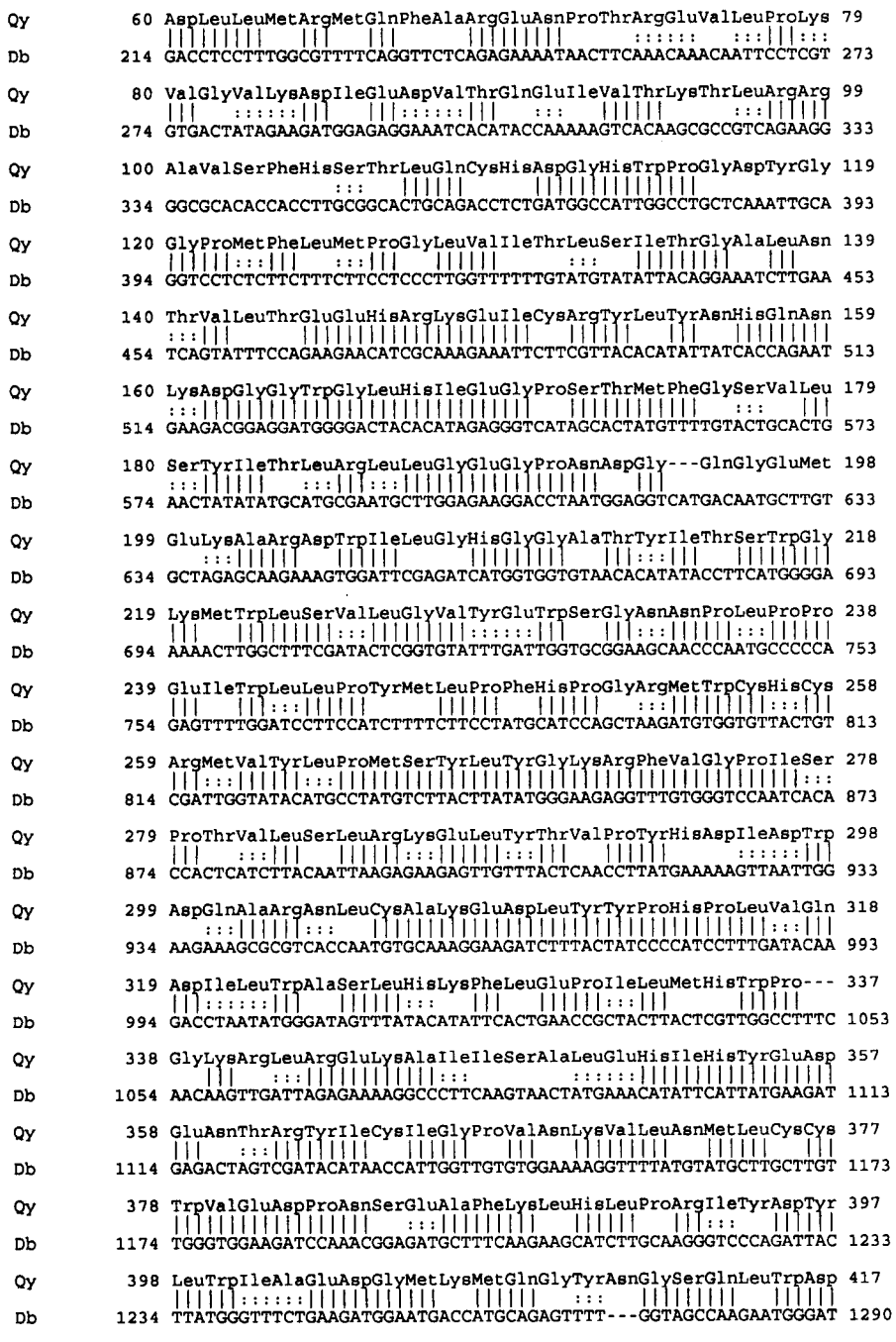
Qy        176 GlySerValLeuSerTyrIleThrLeuArgLeuLeuGlyGluGlyProAsnAspGlyGln 195
                 :::||||||:::|||:::||||||||||||||||:::      :::   |||
Db        370 TGTACTGTCTTGACTTATGTTACCTTGAGATTACTTGGAGACGAATCAGATGGTGGAGAT 429

Qy        196 GlyGluMetGluLysAlaArgAspTrpIleLeuGlyHisGlyGlyAlaThrTyrIleThr 215
              |||  |||     ||||||:::||||||||  ||||||||||||||:::   |||
Db        430 GGAGCCATGGTGATAGCTCGAAATTGGATTTTAGATCATGGAGGAGCAACTTTCACAACA 489

Qy        216 SerTrpGlyLysMetTrpLeuSer---------ValLeuGlyValTyrGluTrpSerGly 232
              ||||||||||||  |||||||||          ||||||||||||:::::|||||||||
Db        490 TCATGGGGAAAATTTTGGCTTTCAGTACTTAAGGTACTTGGTGTCTTTGACTGGTCTGGT 549

Qy        233 AsnAsnProLeuProProGluIleTrpLeuLeuProTyrMetLeuProPheHis----Pr 251
              ||||||||||||  ||||||:::|||:::|||||||||  ||||||||||||    ||
Db        550 AACAACCCATTGCTTCCAGAATTGTGGATGCTACCTTATTTCCTCCCATTTCATCCAGCC 609

Qy        251 oGlyArgMetTrpCysHisCysArgMetValTyrLeuProMetSerTyrLeuTyrGlyLy 271
```



RESULT 9
ADI45184
ID ADI45184 standard; cDNA; 2093 BP.
XX
AC ADI45184;

```
Db       1234 TTATGGGTTTCTGAAGATGGAATGACCATGCAGAGTTTT---GGTAGCCAAGAATGGGAT 1290

Qy        418 ThrAlaPheAlaValGlnAlaIleIleAlaSerAsnLeuIleGluGluPheGlyProThr 437

Db       1291 GCTGGCTTTGCTGTTCAAGCTTTGCTTGCCACTAACATAATTGAAGAAATTGGTCCTACG 1350

Qy        438 IleArgLysAlaHisThrTyrIleLysAsnSerGlnValLeuGluAspCysProGlyAsp 457

Db       1351 TTTGCAAAAGGACATGATTTCATCAAGAAGTCTCAGGTGAAGGATAATCCTTTTGGAGAT 1410

Qy        458 LeuAsnLysTrpTyrArgHisIleSerLysGlyAlaTrpProPheSerThrGlyAspHis 477

Db       1411 TTTAAAAGTATGCATCGTCATATTTCTAAAGGGTCTTGGACATTCTCTGATCAAGACCAT 1470

Qy        478 GlyTrpProIleSerAspCysThrAlaGluGlyLeuLysAlaValLeuLeuLeuSerLys 497

Db       1471 GGATGGCAAGTTTCTGATTGCACTGCAGAAGGTTTAAAGTGTTGTCTACTTCTATCAATG 1530

Qy        498 IleAlaProGluIleValGlyGluProIleAspValLysArgLeuTyrAspSerValAsn 517

Db       1531 TTGCCACCAGAGATTGTGGGAGAAAAGATGGAACCTGAAAGATTATACGATTCAGTCAAT 1590

Qy        518 ValIleLeuSerLeuGlnAsnGluAspGlyGlyPheAlaThrTyrGluLeuLysArgSer 537

Db       1591 GTCTTGTTGTCGCTTCAGAGTAAAAAAGGTGGTTTAGCAGCATGGGAGCCTGCAGGAGCT 1650

Qy        538 TyrAsnTrpLeuGluIleIleAsnProAlaGluThrPheGlyAspIleValIleAspTyr 557

Db       1651 CAAGAGTGGTTAGAATTACTCAATCCCACAGAATTTTTTGCGGACATTGTAGTTGAACAT 1710

Qy        558 ProTyrValGluCysThrSerAlaAlaIleGlnAlaLeuAlaSerPheArgLysLeuTyr 577

Db       1711 GAATATGTTGAGTGCACTGGATCTGCAATCCAAGCTTTAGTTTTGTTCAAGAAGCTATAT 1770

Qy        578 ProGlyHisArgArgGluGluIleGlnHisCysIleAspLysAlaThrThrPheIleGlu 597

Db       1771 CCAGGACATAGGAAGAAAGAGATAGAAAATTTCATTACCAATGCAGTTCGATTCCTTGAA 1830

Qy        598 LysIleGlnAlaSerAspGlySerTrpTyrGlySerTrpGlyValCysPheThrTyrGly 617

Db       1831 GATACACAAACAGCTGATGGTTCATGGTATGGAAATTGGGGAGTTTGCTTCACTTATGGC 1890

Qy        618 AlaTrpPheGlyValLysGlyLeuIleAlaAlaGlyArgSerPheSerAsnCysSerSer 637

Db       1891 TCTTGGTTTGCACTTGGAGGTCTAGCAGCTGCTGGTAAGACTTACACCAATTGTGCTGCC 1950

Qy        638 IleArgLysAlaCysGluPheLeuLeuSerLysGlnLeuProSerGlyGlyTrpGlyGlu 657

Db       1951 ATTCGCAAAGCCGTTAAATTTCTACTTACAACACAAAGAGAGGACGGTGGATGGGGAGAG 2010

Qy        658 SerTyrLeuSerCysGlnAsnLysValTyrSerAsnLeuGluGlyAsnArgSerHisVal 677

Db       2011 AGTTATCTTTCAAGCCCAAAAAAGATATATGTACCTCTAGAAGGAAGCCGATCAAATGTT 2070

Qy        678 ValAsnThrGlyTrpAlaMetLeuAlaLeuIleAspAlaGlyGlnAlaLysArgAspSer 697

Db       2071 GTACATACAGCATGGGCTCTTATGGGACTAATTCATGCTGGACAGGCGGATAGAGACCCC 2130

Qy        698 GlnProLeuHisArgAlaAlaAlaTyrLeuIleAsnSerGlnLeuGluAspGlyAspPhe 717

Db       2131 ATGCCTCTTCACCGTGCTGCAAAGTTGCTCATTAATTCTCAGTTGGAAGAGGGTGATTGG 2190

Qy        718 ProGlnGlnGluIleMetGlyValPheAsnLysAsnCysMetIleThrTyrAlaAlaTyr 737

Db       2191 CCCCAACAGGAAATCACGGGAGTATTCATGAAAAATTGCATGTTGCATTATCCAATGTAC 2250

Qy        738 ArgAsnIleProIleTrpAlaLeuGlyGluTyrGlnSerGlnVal 753

Db       2251 AGAGATATTTATCCAATGTGGGCTCTAGCTGAATATCGAAGGCGGGTT 2298

RESULT 8
ADF89882
ID   ADF89882 standard; cDNA; 2478 BP.
XX
AC   ADF89882;
XX
DT   26-FEB-2004  (first entry)
XX
DE   G. max beta-amyrin synthase cDNA fragment (clone src3c.pk024.m11).
XX
KW   Oxidosqualene cyclase; triterpene; plant protectant; industrial product;
KW   plant; soybean; gene; ss; beta-amyrin synthase.
XX
OS   Glycine max.
XX
PN   WO2003095615-A2.
XX
PD   20-NOV-2003.
XX
PF   05-MAY-2003; 2003WO-US014218.
XX
PR   09-MAY-2002; 2002US-0379361P.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Mcgonigle B,  Maxwell CA,  Hession AO;
XX
DR   WPI; 2004-012098/01.
XX
PT   New transgenic plants comprising recombinant DNA molecules encoding
PT   oxidosqualene cyclase to lower triterpene level, used as feeds, foods,
PT   beverages and industrial products like agricultural adjuvants or concrete
PT   supplies.
XX
PS   Example 1; SEQ ID NO 4; 39pp; English.
XX
CC   The invention relates to a plant comprising at least one recombinant DNA
CC   molecule comprising a promoter operably linked to at least a portion of
CC   at least one oxidosqualene cyclase gene, the molecule sufficient to
CC   suppress the production of a triterpene or any of its progeny, where the
CC   progeny comprise the molecule. The oxidosqualene cyclase gene catalyzes
CC   the cyclization of 2,3-oxidosqualene to form a triterpene, e.g. beta-
CC   amyrin, lanosterol, lupeol, cycloartenol, alpha-amyrin, isomultiflorenol
CC   or any combination. The plants are useful in generating products which
CC   may be used as feeds, foods, beverages and industrial products like
CC   agricultural adjuvants, concrete supplies, dielectric fluids, dust
CC   suppressants, fuel additives, hydraulic fluids, industrial cleaners,
CC   industrial lubricants, metalworking fluids, odor reduction, paint
CC   strippers, printing inks, printing supplies or saw guide oils. Saponins
CC   may play a defense role against pathogens in plant tissues. The present
CC   sequence represents a cDNA insert from clone src3c.pk024.m11 encoding a
CC   G. max beta-amyrin synthase.
XX
SQ   Sequence 2478 BP; 741 A; 455 C; 574 G; 708 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                6.21e-240      Length:          2478
Score:                    2538.00        Matches:         449
Percent Similarity:       75.79%         Conservative:    124
Best Local Similarity:    59.39%         Mismatches:      179
Query Match:              61.19%         Indels:          4
DB:                       12             Gaps:            4

US-09-419-679-6 (1-757) x ADF89882 (1-2478)

Qy          1 MetTrpLysLeuLysPheAlaGluGlyGlyAsn---ProTrpLeuArgThrLeuAsnAsn 19

Db         34 ATGTGGAGGCTGAAGATAGCAGATGGAGGAAATGATCCATACATATTCAGCACAAACAAT 93

Qy         20 HisValGlyArgGlnValTrpGluPheAspProLysLeuGlySerProGlnAspLeuLeu 39

Db         94 TTCGTTGGGAGGCAGACATGGGAGTTTGATCCTGAAGCAGGCAGTCCAGAGGAACGGGCC 153

Qy         40 GluIleGluLysAlaArgGlnAsnPheHisAspAsnArgPheThrHisLysHisSerAla 59

Db        154 CAGGTTGAAGCAGCTCGTCAGCATTTCTACCACAACCGCTTCAAGGTCAAGCCCTGCGCT 213
```

```
KW   miso; tempeh; beta-amyrin synthase; ss.
XX
OS   Glycine max.
XX
FH   Key             Location/Qualifiers
FT   CDS             34. .2322
FT                   /*tag= a
FT                   /product= "Soybean beta-amyrin synthase protein"
XX
PN   WO200166773-A2.
XX
PD   13-SEP-2001.
XX
PF   09-MAR-2001; 2001WO-US007611.
XX
PR   09-MAR-2000; 2000US-0188054P.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Bryan GT,  Mcgonigle B,  Maxwell CA,  Potter SM,  Hwang D;
XX
DR   WPI; 2001-565594/63.
DR   P-PSDB; AAE09795.
XX
PT   New nucleic acid fragment encoding oxidosqualene cyclase, useful as DNA
PT   hybridization probe, polymerase chain reaction primer and for creating
PT   transgenic plants with ce1ls having higher or lower of oxidosqualene
PT   cyclase.
XX
PS   Claim 2; Page 44-45; 59pp; English.
XX
CC   The invention relates to oxidosqualene cyclase proteins and their
CC   corresponding cDNA molecules. Oxidosqualene cyclase genes are used as
CC   hybridisation probes to screen libraries from any desired plant.
CC   Oxidosqualene cyclase DNA are used to create transgenic plants with cells
CC   having higher or lower level of oxidosqualene cyclase. These changes in
CC   the plant seed are useful to improve the seed nutritional value, and in
CC   the plant leaf may aid in insect tolerance. Oxidosqualene cyclase are
CC   produced in heterologous host cells and are used to prepare antibodies.
CC   The antibodies are useful for detecting oxidosqualene cyclase in situ in
CC   cells or in vitro in cell extracts. The invention is useful for altering
CC   level of expression of oxidosqualene cyclase (beta-amyrin synthase)
CC   protein in a plant cell. Soybean grain produced from a transgenic plant
CC   is useful for preparing food products such as soy protein product,
CC   soybean meal, soy flour, soy protein concentrate, soy milk, a dietary
CC   supplement, beans, nuts, tofu, natto, miso and tempeh. The formed soy
CC   protein product having increased saponin levels is useful for lowering
CC   cholesterol in a person suspected of having increased cholesterol, and
CC   for decreasing risk of colon cancer in a person suspected of having risk
CC   of cancer. The present cDNA sequence encodes soybean beta-amyrin synthase
CC   protein
XX
SQ   Sequence 2478 BP; 741 A; 455 C; 574 G; 708 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                6.21e-240      Length:          2478
Score:                    2538.00        Matches:         449
Percent Similarity:       75.79%         Conservative:    124
Best Local Similarity:    59.39%         Mismatches:      179
Query Match:              61.19%         Indels:          4
DB:                       4              Gaps:            4

US-09-419-679-6 (1-757) x AAD16894 (1-2478)

Qy          1 MetTrpLysLeuLysPheAlaGluGlyGlyAsn---ProTrpLeuArgThrLeuAsnAsn 19
              ||||||:::||||||   |||:::|||||||||   |||:::::   :::   ||||||
Db         34 ATGTGGAGGCTGAAGATAGCAGATGGAGGAAATGATCCATACATATTCAGCACAAACAAT 93

Qy         20 HisValGlyArgGlnValTrpGluPheAspProLysLeuGlySerProGlnAspLeuLeu 39
                ||||||||||||   |||||||||||||||:::   |||||||||::::::
Db         94 TTCGTTGGGAGGCAGACATGGGAGTTTGATCCTGAAGCAGGCAGTCCAGAGGAACGGGCC 153

Qy         40 GluIleGluLysAlaArgGlnAsnPheHisAspAsnArgPheThrHisLysHisSerAla 59
                                                                          ::::::|||   |||||||||:::|||:::   |||||||||       |||      |||
Db        154 CAGGTTGAAGCAGCTCGTCAGCATTTCTACCACAACCGCTTCAAGGTCAAGCCCTGCGCT 213

Qy         60 AspLeuLeuMetArgMetGlnPheAlaArgGluAsnProThrArgGluValLeuProLys 79
              |||||||||   |||   |||   |||||||||       ::::::   :::|||:::
Db        214 GACCTCCTTTGGCGTTTTCAGGTTCTCAGAGAAAATAACTTCAAACAAACAATTCCTCGT 273

Qy         80 ValGlyValLysAspIleGluAspValThrGlnGluIleValThrLysThrLeuArgArg 99
              |||   :::::::|||   |||:::::::|||   :::   ||||||       :::|||||||
Db        274 GTGACTATAGAAGATGGAGAGGAAATCACATACCAAAAAGTCACAAGCGCCGTCAGAAGG 333

Qy        100 AlaValSerPheHisSerThrLeuGlnCysHisAspGlyHisTrpProGlyAspTyrGly 119
                           :::   ||||||       ||||||||||||||
Db        334 GGCGCACACCACCTTGCGGCACTGCAGACCTCTGATGGCCATTGGCCTGCTCAAATTGCA 393

Qy        120 GlyProMetPheLeuMetProGlyLeuValIleThrLeuSerIleThrGlyAlaLeuAsn 139
              ||||||:::|||   :::|||   ||||||       :::   |||||||||   |||
Db        394 GGTCCTCTCTTCTTTCTTCCTCCCTTGGTTTTTTGTATGTATATTACAGGAAATCTTGAA 453

Qy        140 ThrValLeuThrGluGluHisArgLysGluIleCysArgTyrLeuTyrAsnHisGlnAsn 159
              :::|||       ||||||||||||||||||||||   ||||||   |||   |||||||||
Db        454 TCAGTATTTCCAGAAGAACATCGCAAAGAAATTCTTCGTTACACATATTATCACCAGAAT 513

Qy        160 LysAspGlyGlyTrpGlyLeuHisIleGluGlyProSerThrMetPheGlySerValLeu 179
              :::||||||||||||||||||||||||||||||   ||||||||||||   :::   |||
Db        514 GAAGACGGAGGATGGGGACTACACATAGAGGGTCATAGCACTATGTTTTGTACTGCACTG 573

Qy        180 SerTyrIleThrLeuArgLeuLeuGlyGluGlyProAsnAspGly---GlnGlyGluMet 198
              :::||||||   :::|||:::||||||||||||||||||   |||
Db        574 AACTATATATGCATGCGAATGCTTGGAGAAGGACCTAATGGAGGTCATGACAATGCTTGT 633

Qy        199 GluLysAlaArgAspTrpIleLeuGlyHisGlyGlyAlaThrTyrIleThrSerTrpGly 218
                 :::||||||   ||||||       ||||||||   |||:::|||   |||||||||
Db        634 GCTAGAGCAAGAAAGTGGATTCGAGATCATGGTGGTGTAACACATATACCTTCATGGGGA 693

Qy        219 LysMetTrpLeuSerValLeuGlyValTyrGluTrpSerGlyAsnAsnProLeuProPro 238
              |||   |||||||||:::|||||||||::::::|||   |||:::||||||:::||||||
Db        694 AAAACTTGGCTTTCGATACTCGGTGTATTTGATTGGTGCGGAAGCAACCCAATGCCCCCA 753

Qy        239 GluIleTrpLeuLeuProTyrMetLeuProPheHisProGlyArgMetTrpCysHisCys 258
              |||   |||:::||||||      ||||||   |||||||   :::||||||||||:::|||
Db        754 GAGTTTTGGATCCTTCCATCTTTTCTTCCTATGCATCCAGCTAAGATGTGGTGTTACTGT 813

Qy        259 ArgMetValTyrLeuProMetSerTyrLeuTyrGlyLysArgPheValGlyProIleSer 278
              |||:::||||||:::||||||||||||||||||||||||||||||||||||||||||:::
Db        814 CGATTGGTATACATGCCTATGTCTTACTTATATGGGAAGAGGTTTGTGGGTCCAATCACA 873

Qy        279 ProThrValLeuSerLeuArgLysGluLeuTyrThrValProTyrHisAspIleAspTrp 298
              |||   :::|||   ||||||:::||||||:::|||   ||||||       ::::::|||
Db        874 CCACTCATCTTACAATTAAGAGAAGAGTTGTTTACTCAACCTTATGAAAAAGTTAATTGG 933

Qy        299 AspGlnAlaArgAsnLeuCysAlaLysGluAspLeuTyrTyrProHisProLeuValGln 318
                 :::||||||:::   ||||||||||||||||||||||||||||||||||||:::|||
Db        934 AAGAAAGCGCGTCACCAATGTGCAAAGGAAGATCTTTACTATCCCCATCCTTTGATACAA 993

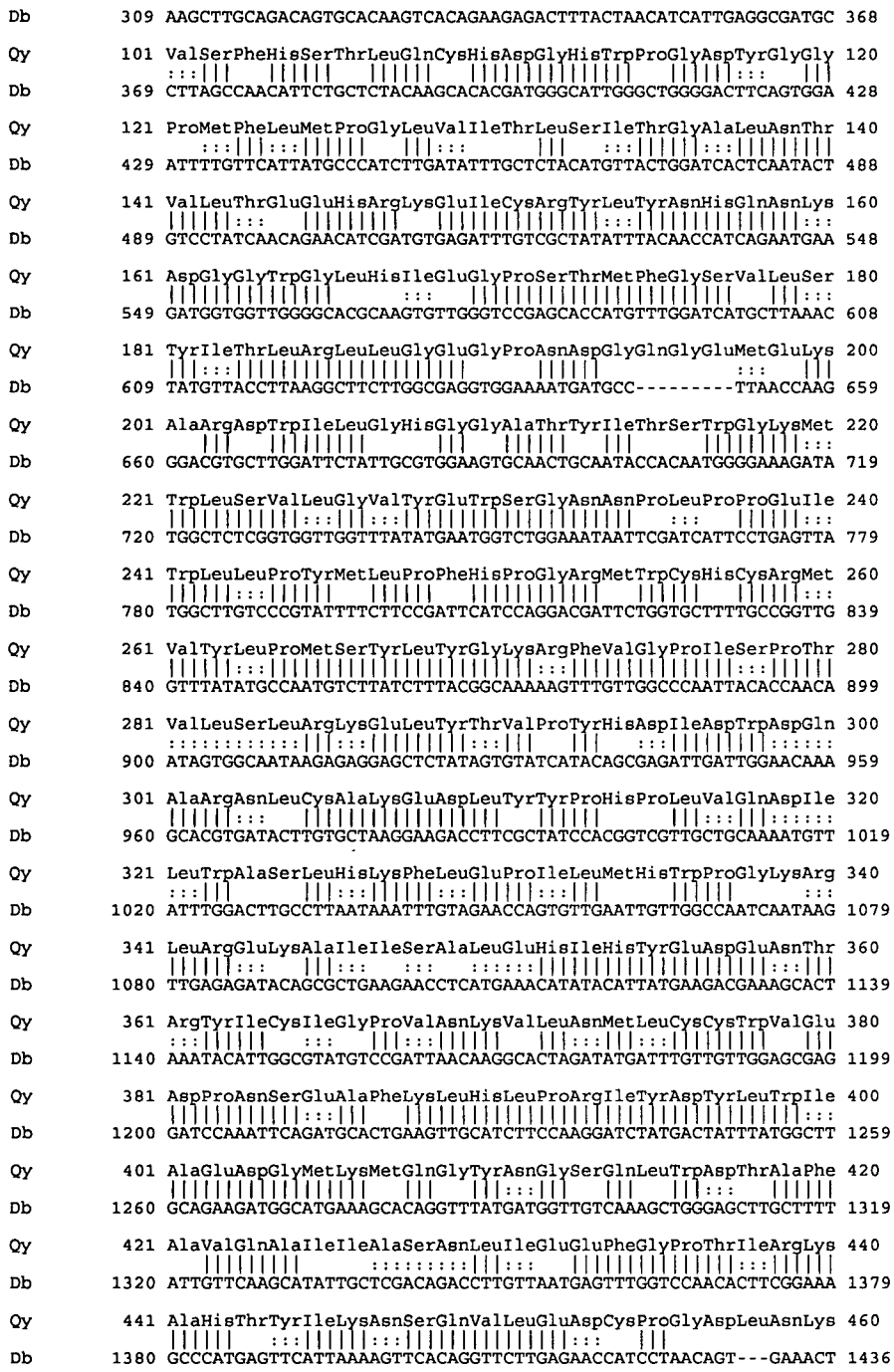
Qy        319 AspIleLeuTrpAlaSerLeuHisLysPheLeuGluProIleLeuMetHisTrpPro--- 337
              |||:::::|||   ||||||:::   |||   |||||||:::|||       ||||||
Db        994 GACCTAATATGGGATAGTTTATACATATTCACTGAACCGCTACTTACTCGTTGGCCTTTC 1053

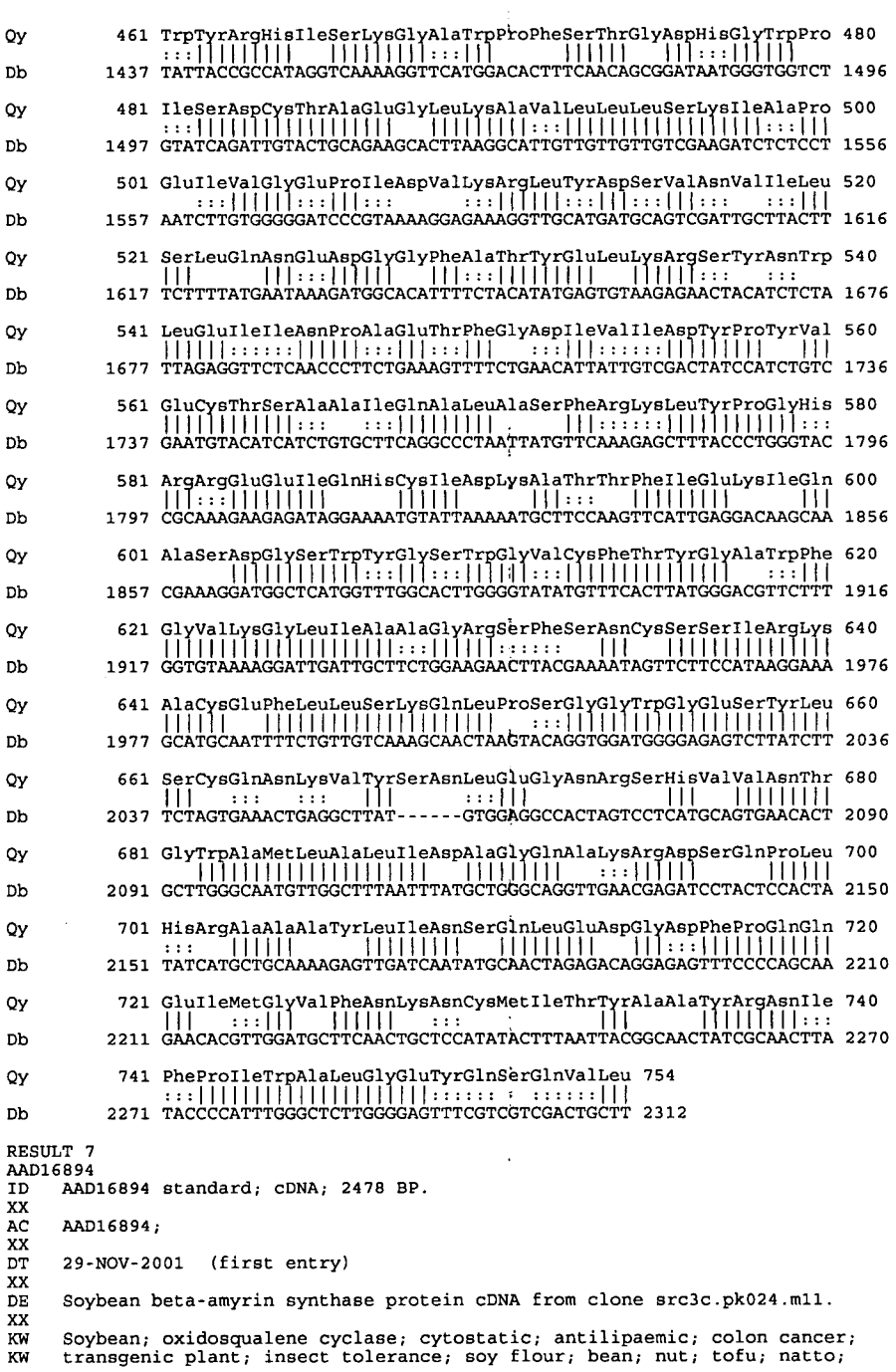
Qy        338 GlyLysArgLeuArgGluLysAlaIleIleSerAlaLeuGluHisIleHisTyrGluAsp 357
                   |||   :::|||||||||||||:::         ::::::||||||||||||||||||
Db       1054 AACAAGTTGATTAGAGAAAAGGCCCTTCAAGTAACTATGAAACATATTCATTATGAAGAT 1113

Qy        358 GluAsnThrArgTyrIleCysIleGlyProValAsnLysValLeuAsnMetLeuCysCys 377
              |||   :::||||||||||   ||||||   |||   |||||||||   ||||||   |||
Db       1114 GAGACTAGTCGATACATAACCATTGGTTGTGTGGAAAAGGTTTTATGTATGCTTGCTTGT 1173

Qy        378 TrpValGluAspProAsnSerGluAlaPheLysLeuHisLeuProArgIleTyrAspTyr 397
              |||||||||||||||||   :::|||||||||   ||||||   |||:::   ||||||
Db       1174 TGGGTGGAAGATCCAAACGGAGATGCTTTCAAGAAGCATCTTGCAAGGGTCCCAGATTAC 1233

Qy        398 LeuTrpIleAlaGluAspGlyMetLysMetGlnGlyTyrAsnGlySerGlnLeuTrpAsp 417
              ||||||::::::|||||||||||||   ||||||   :::   |||||||||   ||||||
```





RESULT 7
AAD16894
ID AAD16894 standard; cDNA; 2478 BP.
XX
AC AAD16894;
XX
DT 29-NOV-2001 (first entry)
XX
DE Soybean beta-amyrin synthase protein cDNA from clone src3c.pk024.m11.
XX
KW Soybean; oxidosqualene cyclase; cytostatic; antilipaemic; colon cancer;
KW transgenic plant; insect tolerance; soy flour; bean; nut; tofu; natto;

```
Qy        514 AspSerValAsnValIleLeuSerLeuGlnAsnGluAspGlyGlyPheAlaThrTyrGlu 533
              |||:::|||:::   :::::::|||       |||::::::||||||||||||||||||||
Db       1383 GATGCTGTCAGTTGTCTGATGTCTTATATGAATGATAATGGCGGTTTCGCGACATATGAA 1442

Qy        534 LeuLysArgSerTyrAsnTrpLeuGluIleIleAsnProAlaGluThrPheGlyAspIle 553
              |||   |||||||||   ||||||||||:::||||||||||||||||||||||||||||
Db       1443 CTCACAAGATCTTATCCCTGGTTGGAGCTGATCAATCCCGCTGAGACCTTTGGGGATATT 1502

Qy        554 ValIleAspTyrProTyrValGluCysThrSerAlaAlaIleGlnAlaLeuAlaSerPhe 573
              |||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db       1503 GTGATTGATTACCCGTATGTTGAATGTACATCAGCAGCAATTCAGGCCCTGACATCATTC 1562

Qy        574 ArgLysLeuTyrProGlyHisArgArgGluGluIleGlnHisCysIleAspLysAlaThr 593
              :::|||||||||||||||||||||||:::|||:::   :::||||||   ||||||:::
Db       1563 AAAAAACTATACCCTGGGCACCGCAGGAAAGAGGTGGATAACTGTATCAGCAAAGCTTCC 1622

Qy        594 ThrPheIleGluLysIleGlnAlaSerAspGlySerTrpTyrGlySerTrpGlyValCys 613
                 |||||||||   ||||||   ||||||||||||||||||||||||||   ||||||
Db       1623 AATTTCATCGAGAGTATTCAGAAAAGCGATGGTTCATGGTATGGCTCTTGGGCCGTCTGT 1682

Qy        614 PheThrTyrGlyAlaTrpPheGlyValLysGlyLeuIleAlaAlaGlyArgSerPheSer 633
              |||||||||||   |||||||||||||||||||||||||||||||||:::|||
Db       1683 TTCACATACGGCACTTGGTTTGGTGTGAAGGGACTAATTGCTGCTGGTAGAACATTTGAG 1742

Qy        634 AsnCysSerSerIleArgLysAlaCysGluPheLeuLeuSerLysGlnLeuProSerGly 653
              |||      :::|||||||||||||||:::|||||||||||||||:::||||||||||||
Db       1743 AACAGTCCTGCAATTAGAAAGGCATGCGACTTTCTGTTGTCAAAAGAACTTCCTTCCGGT 1802

Qy        654 GlyTrpGlyGluSerTyrLeuSerCysGlnAsnLysValTyrSerAsnLeuGluGlyAsn 673
              |||||||||||||||||||||||   |||::::::|||||:::||||||:::||||||
Db       1803 GGTTGGGGAGAAAGCTATTTGTCATCTCAAGACCAGGTTTACACCAATCTCAAAGGCAAC 1862

Qy        674 ArgSerHisValValAsnThrGlyTrpAlaMetLeuAlaLeuIleAspAlaGlyGlnAla 693
              |||   |||   |||||||||   ||||||||||||||||||||||||||||||||||
Db       1863 CGGCCTCATGCGGTGAACACTAGTTGGGCCATGCTGGCGCTGATTGATGCGGGCCAGGCC 1922

Qy        694 LysArgAspSerGlnProLeuHisArgAlaAlaAlaTyrLeuIleAsnSerGlnLeuGlu 713
              :::|||||||      ||||||||||||||||||      |||||||||   |||   |||
Db       1923 GAGAGAGATCCAACGCCTCTACACCGAGCAGCAAAGGTTTTGATCAACTTACAATCAGAG 1982

Qy        714 AspGlyAspPheProGlnGlnGluIleMetGlyValPheAsnLysAsnCysMetIleThr 733
              ||||||:::||||||||||||||||||:::||||||||||||||||||||||||||:::
Db       1983 GACGGAGAATTTCCTCAGCAAGAGATCATAGGAGTGTTCAACAAGAACTGCATGATAAGC 2042

Qy        734 TyrAlaAlaTyrArgAsnIlePheProIleTrpAlaLeuGlyGluTyrGlnSerGlnVal 753
              |||:::   ||||||||||||||||||||||||||||||||||||:::   :::|||
Db       2043 TACTCCCAGTACAGGAACATCTTCCCGATTTGGGCTCTGGGTGAGTACCGGTGTCGAGTC 2102

Qy        754 Leu 754
              |||
Db       2103 TTG 2105
```

```
RESULT 6
AAD16896
ID   AAD16896 standard; cDNA; 2538 BP.
XX
AC   AAD16896;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Wheat oxidosqualene cyclase protein cDNA from clone wdk1c.pk010.o10.
XX
KW   Wheat; oxidosqualene cyclase; cytostatic; antilipaemic; colon cancer;
KW   transgenic plant; insect tolerance; soy flour; bean; nut; tofu; natto;
KW   miso; tempeh; ss.
XX
OS   Triticum aestivum.
XX
FH   Key             Location/Qualifiers
FT   CDS             69. .2324
FT                   /*tag= a
FT                   /product= "Wheat oxidosqualene cyclase protein"
XX
PN   WO200166773-A2.
XX
PD   13-SEP-2001.
XX
PF   09-MAR-2001; 2001WO-US007611.
XX
PR   09-MAR-2000; 2000US-0188054P.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Bryan GT,  Mcgonigle B,  Maxwell CA,  Potter SM,  Hwang D;
XX
DR   WPI; 2001-565594/63.
DR   P-PSDB; AAE09797.
XX
PT   New nucleic acid fragment encoding oxidosqualene cyclase, useful as DNA
PT   hybridization probe, polymerase chain reaction primer and for creating
PT   transgenic plants with cells having higher or lower of oxidosqualene
PT   cyclase.
XX
PS   Claim 2; Page 50-51; 59pp; English.
XX
CC   The invention relates to oxidosqualene cyclase proteins and their
CC   corresponding cDNA molecules. Oxidosqualene cyclase genes are used as
CC   hybridisation probes to screen libraries from any desired plant.
CC   Oxidosqualene cyclase DNA are used to create transgenic plants with cells
CC   having higher or lower level of oxidosqualene cyclase. These changes in
CC   the plant seed are useful to improve the seed nutritional value, and in
CC   the plant leaf may aid in insect tolerance. Oxidosqualene cyclase are
CC   produced in heterologous host cells and are used to prepare antibodies.
CC   The antibodies are useful for detecting oxidosqualene cyclase in situ in
CC   cells or in vitro in cell extracts. The invention is useful for altering
CC   level of expression of oxidosqualene cyclase (beta-amyrin synthase)
CC   protein in a plant cell. Soybean grain produced from a transgenic plant
CC   is useful for preparing food products such as soy protein product,
CC   soybean meal, soy flour, soy protein concentrate, soy milk, a dietary
CC   supplement, beans, nuts, tofu, natto, miso and tempeh. The formed soy
CC   protein product having increased saponin levels is useful for lowering
CC   cholesterol in a person suspected of having increased cholesterol, and
CC   for decreasing risk of colon cancer in a person suspected of having risk
CC   of cancer. The present cDNA sequence encodes wheat oxidosqualene cyclase
CC   protein
XX
SQ   Sequence 2538 BP; 727 A; 486 C; 613 G; 712 T; 0 U; 0 Other;
```

Alignment Scores:			
Pred. No.:	3.34e-241	Length:	2538
Score:	2551.00	Matches:	446
Percent Similarity:	77.59%	Conservative:	139
Best Local Similarity:	59.15%	Mismatches:	163
Query Match:	61.50%	Indels:	6
DB:	4	Gaps:	3

US-09-419-679-6 (1-757) x AAD16896 (1-2538)

```
Qy          1 MetTrpLysLeuLysPheAlaGluGlyGlyAsnProTrpLeuArgThrLeuAsnAsnHis 20
              ||||||:::||||||   :::||||||||||  ||||||||||||:::::||||||
Db         69 ATGTGGAGGCTCAAGGTGTCCGAGGGTGGCGGCCCGTGGCTGCGGTCGGTGAACAACTTC 128

Qy         21 ValGlyArgGlnValTrpGluPheAspProLysLeuGlySerProGlnAspLeuLeuGlu 40
              :::|||||||   ||||||||||||||||||      |||:::|||::::::      |||
Db        129 CTCGGCAGGGCAGTGTGGGAGTTCGACCCCGACTACGGCACACCGGAGGAGCGCGCCGAG 188

Qy         41 IleGluLysAlaArgGlnAsnPheHisAspAsnArgPheThrHisLysHisSerAlaAsp 60
              :::::::::   |||:::   |||   |||   ||||||      |||   |||   |||
Db        189 GTGAAGAGGGTGCGCCGGGAGTTCACCGACCGCCGTTTCGAGAAAAAGGAGTCGCAGGAT 248

Qy         61 LeuLeuMetArgMetGlnPheAlaArgGluAsnProThrArgGluValLeuProLysVal 80
              ||||||||||||||||||:::|||:::|||          :::      ||||||   :::
Db        249 CTTCTCATGCGCATGCAGTATGCAAAAGAAAAGCATCTTCAGGTGGACCTTCCAGCCATC 308

Qy         81 GlyValLysAspIleGluAspValThrGlnGluIleValThrLysThrLeuArgArgAla 100
                 :::   |||          ||||||:::|||   :::      :::||||||||||
```

```
PA   (FAMO/) FAMODU O O.
PA   (HITZ/) HITZ W D.
PA   (KINN/) KINNEY A J.
PA   (RAFA/) RAFALSKI J A.
PA   (MCGO/) MCGONIGLE B.
PA   (LOHM/) LOHMAN K N.
XX
PI   Falco SC,  Famodu OO,  Hitz WD,  Kinney AJ,  Rafalski JA;
PI   Mcgonigle B,  Lohman KN;
XX
DR   WPI; 2004-479684/45.
DR   P-PSDB; ADQ95184.
XX
PT   New composition comprising enzymes involved in squalene metabolism,
PT   useful in altering squalene metabolism in plants and seeds.
XX
PS   Claim 2; SEQ ID NO 1; 33pp; English.
XX
CC   The invention describes a composition comprising an isolated
CC   polynucleotide comprising a nucleotide sequence encoding a first
CC   polypeptide of at least 200 amino acids. A composition comprising an
CC   isolated polynucleotide comprising a nucleotide sequence encoding a first
CC   polypeptide of at least 200 amino acids that has at least 90% identity
CC   based on the Clustal method of alignment when compared to a polypeptide
CC   comprising a sequence of 701, 541, 757, 345, 386, 330, 523, or 467 amino
CC   acids (EVEN SEQ ID NOS: 2-16), or an isolated polynucleotide comprising
CC   the complement of the nucleotide sequence, or comprising a polypeptide of
CC   at least 200 amino acids that has at least 90% identity based on the
CC   Clustal method of alignment when compared to a polypeptide of EVEN SEQ ID
CC   NOS: 2-16. Also described are: a chimeric gene comprising the isolated
CC   polynucleotide described above operably linked to suitable regulatory
CC   sequences; an isolated host cell comprising the chimeric gene of (1) or
CC   the isolated polynucleotide described above; a virus comprising the
CC   isolated polynucleotide described above; selecting an isolated
CC   polynucleotide that affects the level of expression of a polypeptide in a
CC   host or plant cell; obtaining a nucleic acid fragment encoding a
CC   polypeptide; and evaluating at least one compound for its ability to
CC   inhibit the activity of a polypeptide. The composition, methods,
CC   polynucleotides and polypeptides are useful in altering squalene
CC   metabolism in plants and seeds. This sequence encodes corn cycloartenol
CC   synthase, a member of the squalene metabolism pathway.
XX
SQ   Sequence 2558 BP; 706 A; 532 C; 636 G; 684 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                6.62e-298      Length:          2558
Score:                    3125.00        Matches:         545
Percent Similarity:       88.87%.        Conservative:    78
Best Local Similarity:    77.75%         Mismatches:      78
Query Match:              75.34%         Indels:          0
DB:                       12             Gaps:            0

US-09-419-679-6 (1-757) x ADQ95183 (1-2558)

Qy          54 ThrHisLysHisSerAlaAspLeuLeuMetArgMetGlnPheAlaArgGluAsnProThr 73
               |||  ||||||||||||||||||||||||||||||:::|||||||||:::|||||||
Db           3 ACGAGCAAGCACAGCGCCGACCTCCTCATGCGCATCCAGTTCGCCAAAGAAAACTCGATT 62

Qy          74 ArgGluValLeuProLysValGlyValLysAspIleGluAspValThrGlnGluIleVal 93
                     ||||||   :::   :::   :::   ||||||||||||:::|||   |||
Db          63 GAGCTTCACCTTCCAGGCATCAAGCTCGGTGAGCATGAAGATGTGACCGAGGAAGCTGTG 122

Qy          94 ThrLysThrLeuArgArgAlaValSerPheHisSerThrLeuGlnCysHisAspGlyHis 113
                    ||||||:::|||||||:::|||      ||||||||||||   ||||||||||||
Db         123 TTGACTACATTGAAAAGGGCAATCAGCCGTTTCTCTACTCTCCAGGCACATGATGGACAC 182

Qy         114 TrpProGlyAspTyrGlyGlyProMetPheLeuMetProGlyLeuValIleThrLeuSer 133
               ||||||||||||||||||||||||||||||||||||||||||||:::||||||||||
Db         183 TGGCCTGGGGATTATGGTGGTCCTATGTTCCTTATGCCAGGCTTGATCATAACATTGTAT 242

Qy         134 IleThrGlyAlaLeuAsnThrValLeuThrGluGluHisArgLysGluIleCysArgTyr 153
               :::|||||||||||||||||||||||:::   ||||||:::||||||||   ||||||
Db         243 GTGACTGGAGCACTAAACACTGTCTTGTCATTGGAACACCAGAAGGAGATCCGCCGGTAT 302

Qy         154 LeuTyrAsnHisGlnAsnLysAspGlyGlyTrpGlyLeuHisIleGluGlyProSerThr 173
               ||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||
Db         303 CTTTATAATCACCAGAATGAAGATGGCGGCTGGGGCTTGCACATTGAGGGTCCAAGCACC 362

Qy         174 MetPheGlySerValLeuSerTyrIleThrLeuArgLeuLeuGlyGluGlyProAsnAsp 193
               ||||||||||||   |||:::|||:::   |||||||||||||||||||||||:::
Db         363 ATGTTCGGCTCAGCACTGACCTATGTTATTTTGAGATTGCTTGGAGAGGGACCAGATAGT 422

Qy         194 GlyGlnGlyGluMetGluLysAlaArgAspTrpIleLeuGlyHisGlyGlyAlaThrTyr 213
               |||   |||   |||||||||   |||:::|||||||||   ||||||||||||||||||
Db         423 GGAGATGGAGCCATGGAGAAAGGTCGAAACTGGATATTAGACCATGGTGGAGCAACATAT 482

Qy         214 IleThrSerTrpGlyLysMetTrpLeuSerValLeuGlyValTyrGluTrpSerGlyAsn 233
               ||||||||||||||||||   ||||||||||||||||||||:::||||||||||||||||
Db         483 ATAACATCGTGGGGGAAGTTTTGGCTTTCGGTACTAGGTGTATTTGAATGGTCTGGTAAC 542

Qy         234 AsnProLeuProProGluIleTrpLeuLeuProTyrMetLeuProPheHisProGlyArg 253
               ||||||:::|||||||||:::|||||||||||||||:::||||||||||||||||||||
Db         543 AACCCGGTGCCACCAGAAGTATGGCTACTGCCATATCTCCTCCCATTTCACCCAGGGAGG 602

Qy         254 MetTrpCysHisCysArgMetValTyrLeuProMetSerTyrLeuTyrGlyLysArgPhe 273
               ||||||||||||||||||||||||||||||||||   |||:::|||||||||||||||
Db         603 ATGTGGTGTCATTGTCGAATGGTGTATTTGCCAATGTGCTACATTTATGGGAAGAGGTTT 662

Qy         274 ValGlyProIleSerProThrValLeuSerLeuArgLysGluLeuTyrThrValProTyr 293
               ||||||   |||:::|||   :::|||   ||||||||||||||:::      ||||||
Db         663 GTTGGCCGAATCACACCACTTCTGTTGGAATTAAGGAAGGAACTTTTCAAAGACCCTTAC 722

Qy         294 HisAspIleAspTrpAspGlnAlaArgAsnLeuCysAlaLysGluAspLeuTyrTyrPro 313
               ||||||||||:::||||||||||||||||||||||||||||||||||||||||||||||
Db         723 AGCAAGATTGATTGGGACAAGGCCCGCAACCTATGTGCCAAGGAAGATCTGTACTACCCA 782

Qy         314 HisProLeuValGlnAspIleLeuTrpAlaSerLeuHisLysPheLeuGluProIleLeu 333
               ||||||   |||||||||:::|||||||||:::||||||||||||:::||||||:::::
Db         783 CACCCATTCGTTCAAGATGTGTTGTGGGCCACTCTCCATAAATTCGTTGAACCAGTTATG 842

Qy         334 MetHisTrpProGlyLysArgLeuArgGluLysAlaIleIleSerAlaLeuGluHisIle 353
               ||||||||||||||   :::|||||||||||||||:::   :::   ::::::|||:::
Db         843 ATGCATTGGCCTGGCAGCAAATTGAGGGAGAAAGCTCTGGAAACAGTCATGCAACATGTT 902

Qy         354 HisTyrGluAspGluAsnThrArgTyrIleCysIleGlyProValAsnLysValLeuAsn 373
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         903 CATTATGAAGATGAGAACACTCGTTATATTTGCATTGGTCCTGTAAACAAGGTATTGAAT 962

Qy         374 MetLeuCysCysTrpValGluAspProAsnSerGluAlaPheLysLeuHisLeuProArg 393
               ||||||   ||||||:::|||||||||||||||||||||||||||||||||:::||||||
Db         963 ATGCTTGCTTGCTGGATTGAAGATCCAAACTCGGAGGCCTTCAAACTTCATATCCCACGA 1022

Qy         394 IleTyrAspTyrLeuTrpIleAlaGluAspGlyMetLysMetGlnGlyTyrAsnGlySer 413
               :::|||||||||||||:::|||||||||||||||||||||||||||||||||||||||||
Db        1023 GTCTATGATTACTTGTGGCTTGCTGAAGATGGCATGAAGATGCAGGGTTATAATGGCAGC 1082

Qy         414 GlnLeuTrpAspThrAlaPheAlaValGlnAlaIleIleAlaSerAsnLeuIleGluGlu 433
               ||||||||||||||||||   ||||||||||:::|||:::||||||||||||||||||
Db        1083 CAACTTTGGGATACAGCTTTCACAGTTCAAGCCATTGTGGCTACCAACCTTATTGAAGAG 1142

Qy         434 PheGlyProThrIleArgLysAlaHisThrTyrIleLysAsnSerGlnValLeuGluAsp 453
               ||||||||||||:::::   ||||||   ||||||||||||||||||||||||:::|||
Db        1143 TTTGGTCCTACCCTTAAACTAGCACACAACTATATCAAGAATTCACAGGTTCTTGATGAC 1202

Qy         454 CysProGlyAspLeuAsnLysTrpTyrArgHisIleSerLysGlyAlaTrpProPheSer 473
               ||||||||||||||||||   ||||||||||||   ||||||||||||||||||||||||
Db        1203 TGCCCTGGTGATCTGAATGACTGGTACCGCCACACATCTAAAGGTGCATGGCCATTCTCA 1262

Qy         474 ThrGlyAspHisGlyTrpProIleSerAspCysThrAlaGluGlyLeuLysAlaValLeu 493
               |||   ||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db        1263 ACTGCTGATCATGGTTGGCCTATATCTGATTGCACTGCTGAAGGACTAAAGGCTTCATTA 1322

Qy         494 LeuLeuSerLysIleAlaProGluIleValGlyGluProIleAspValLysArgLeuTyr 513
               |||||||||:::|||:::|||:::||||||||||||||:::::      |||   |||
Db        1323 TTGTTATCAAGGATCTCTCCCAAAATTGTTGGTGAACCGATGGAAGCTAATAGATTTTAT 1382
```

```
Db        423 GGAGATGGAGCCATGGAGAAAGGTCGAAACTGGATATTAGACCATGGTGGAGCAACATAT 482
Qy        214 IleThrSerTrpGlyLysMetTrpLeuSerValLeuGlyValTyrGluTrpSerGlyAsn 233
Db        483 ATAACATCGTGGGGGAAGTTTTGGCTTTCGGTACTAGGTGTATTTGAATGGTCTGGTAAC 542
Qy        234 AsnProLeuProProGluIleTrpLeuLeuProTyrMetLeuProPheHisProGlyArg 253
Db        543 AACCCGGTGCCACCAGAAGTATGGCTACTGCCATATCTCCTCCCATTTCACCCAGGGAGG 602
Qy        254 MetTrpCysHisCysArgMetValTyrLeuProMetSerTyrLeuTyrGlyLysArgPhe 273
Db        603 ATGTGGTGTCATTGTCGAATGGTGTATTTGCCAATGTGCTACATTTATGGGAAGAGGTTT 662
Qy        274 ValGlyProIleSerProThrValLeuSerLeuArgLysGluLeuTyrThrValProTyr 293
Db        663 GTTGGCCGAATCACACCACTTCTGTTGGAATTAAGGAAGGAACTTTTCAAAGACCCTTAC 722
Qy        294 HisAspIleAspTrpAspGlnAlaArgAsnLeuCysAlaLysGluAspLeuTyrTyrPro 313
Db        723 AGCAAGATTGATTGGGACAAGGCCCGCAACCTATGTGCCAAGGAAGATCTGTACTACCCA 782
Qy        314 HisProLeuValGlnAspIleLeuTrpAlaSerLeuHisLysPheLeuGluProIleLeu 333
Db        783 CACCCATTCGTTCAAGATGTGTTGTGGGCCACTCTCCATAAATTCGTTGAACCAGTTATG 842
Qy        334 MetHisTrpProGlyLysArgLeuArgGluLysAlaIleIleSerAlaLeuGluHisIle 353
Db        843 ATGCATTGGCCTGGCAGCAAATTGAGGGAGAAAGCTCTGGAAACAGTCATGCAACATGTT 902
Qy        354 HisTyrGluAspGluAsnThrArgTyrIleCysIleGlyProValAsnLysValLeuAsn 373
Db        903 CATTATGAAGATGAGAACACTCGTTATATTTGCATTGGTCCTGTAAACAAGGTATTGAAT 962
Qy        374 MetLeuCysCysTrpValGluAspProAsnSerGluAlaPheLysLeuHisLeuProArg 393
Db        963 ATGCTTGCTTGCTGGATTGAAGATCCAAACTCGGAGGCCTTCAAACTTCATATCCCACGA 1022
Qy        394 IleTyrAspTyrLeuTrpIleAlaGluAspGlyMetLysMetGlnGlyTyrAsnGlySer 413
Db       1023 GTCTATGATTACTTGTGGCTTGCTGAAGATGGCATGAAGATGCAGGGTTATAATGGCAGC 1082
Qy        414 GlnLeuTrpAspThrAlaPheAlaValGlnAlaIleIleAlaSerAsnLeuIleGluGlu 433
Db       1083 CAACTTTGGGATACAGCTTTCACAGTTCAAGCCATTGTGGCTACCAACCTTATTGAAGAG 1142
Qy        434 PheGlyProThrIleArgLysAlaHisThrTyrIleLysAsnSerGlnValLeuGluAsp 453
Db       1143 TTTGGTCCTACCCTTAAACTAGCACACAACTATATCAAGAATTCACAGGTTCTTGATGAC 1202
Qy        454 CysProGlyAspLeuAsnLysTrpTyrArgHisIleSerLysGlyAlaTrpProPheSer 473
Db       1203 TGCCCTGGTGATCTGAATGACTGGTACCGCCACACATCTAAAGGTGCATGGCCATTCTCA 1262
Qy        474 ThrGlyAspHisGlyTrpProIleSerAspCysThrAlaGluGlyLeuLysAlaValLeu 493
Db       1263 ACTGCTGATCATGGTTGGCCTATATCTGATTGCACTGCTGAAGGACTAAAGGCTTCATTA 1322
Qy        494 LeuLeuSerLysIleAlaProGluIleValGlyGluProIleAspValLysArgLeuTyr 513
Db       1323 TTGTTATCAAGGATCTCTCCCAAAATTGTTGGTGAACCGATGGAAGCTAATAGATTTTAT 1382
Qy        514 AspSerValAsnValIleLeuSerLeuGlnAsnGluAspGlyGlyPheAlaThrTyrGlu 533
Db       1383 GATGCTGTCAGTTGTCTGATGTCTTATATGAATGATAATGGCGGTTTCGCGACATATGAA 1442
Qy        534 LeuLysArgSerTyrAsnTrpLeuGluIleIleAsnProAlaGluThrPheGlyAspIle 553
Db       1443 CTCACAAGATCTTATCCCTGGTTGGAGCTGATCAATCCCGCTGAGACCTTTGGGGATATT 1502
Qy        554 ValIleAspTyrProTyrValGluCysThrSerAlaAlaIleGlnAlaLeuAlaSerPhe 573
Db       1503 GTGATTGATTACCCGTATGTTGAATGTACATCAGCAGCAATTCAGGCCCTGACATCATTC 1562
Qy        574 ArgLysLeuTyrProGlyHisArgArgGluGluIleGlnHisCysIleAspLysAlaThr 593
Db       1563 AAAAAACTATACCCTGGGCACCGCAGGAAAGAGGTGGATAACTGTATCAGCAAAGCTTCC 1622
Qy        594 ThrPheIleGluLysIleGlnAlaSerAspGlySerTrpTyrGlySerTrpGlyValCys 613
Db       1623 AATTTCATCGAGAGTATTCAGAAAAGCGATGGTTCATGGTATGGCTCTTGGGCCGTCTGT 1682
Qy        614 PheThrTyrGlyAlaTrpPheGlyValLysGlyLeuIleAlaAlaGlyArgSerPheSer 633
Db       1683 TTCACATACGGCACTTGGTTTGGTGTGAAGGGACTAATTGCTGCTGGTAGAACATTTGAG 1742
Qy        634 AsnCysSerSerIleArgLysAlaCysGluPheLeuLeuSerLysGlnLeuProSerGly 653
Db       1743 AACAGTCCTGCAATTAGAAAGGCATGCGACTTTCTGTTGTCAAAAGAACTTCCTTCCGGT 1802
Qy        654 GlyTrpGlyGluSerTyrLeuSerCysGlnAsnLysValTyrSerAsnLeuGluGlyAsn 673
Db       1803 GGTTGGGGAGAAAGCTATTTGTCATCTCAAGACCAGGTTTACACCAATCTCAAAGGCAAC 1862
Qy        674 ArgSerHisValValAsnThrGlyTrpAlaMetLeuAlaLeuIleAspAlaGlyGlnAla 693
Db       1863 CGGCCTCATGCGGTGAACACTAGTTGGGCCATGCTGGCGCTGATTGATGCGGGCCAGGCC 1922
Qy        694 LysArgAspSerGlnProLeuHisArgAlaAlaAlaTyrLeuIleAsnSerGlnLeuGlu 713
Db       1923 GAGAGAGATCCAACGCCTCTACACCGAGCAGCAAAGGTTTTGATCAACTTACAATCAGAG 1982
Qy        714 AspGlyAspProProGlnGlnGluIleMetGlyValPheAsnLysAsnCysMetIleThr 733
Db       1983 GACGGAGAATTTCCTCAGCAAGAGATCATAGGAGTGTTCAACAAGAACTGCATGATAAGC 2042
Qy        734 TyrAlaAlaTyrArgAsnIleProIleTrpAlaLeuGlyGluTyrGlnSerGlnVal 753
Db       2043 TACTCCCAGTACAGGAACATCTTCCCGATTTGGGCTCTGGGTGAGTACCGGTGTCGAGTC 2102
Qy        754 Leu 754
Db       2103 TTG 2105
```

```
RESULT 5
ADQ95183
ID   ADQ95183 standard; cDNA; 2558 BP.
XX
AC   ADQ95183;
XX
DT   23-SEP-2004  (first entry)
XX
DE   Corn cycloartenol synthase cDNA.
XX
KW   corn; cycloartenol synthase; squalene metabolism; gene; ss.
XX
OS   Zea mays.
XX
FH   Key             Location/Qualifiers
FT   CDS             3. .2105
FT                   /*tag= a
FT                   /product= "Corn cycloartenol synthase"
FT                   /partial
FT                   /note= "No start or stop codon given"
XX
PN   US2004121439-A1.
XX
PD   24-JUN-2004.
XX
PF   27-JUN-2003; 2003US-00607726.
XX
PR   23-OCT-1998;   98US-0105405P.
PR   15-OCT-1999;   99US-00419679.
XX
PA   (FALC/) FALCO S C.
```

```
Db       1501 GCGATTGTTGGTGAACCAATAGATGCAAAACGGTTATATGAAGCTGTTAATGTTATCATT 1560

Qy        521 SerLeuGlnAsnGluAspGlyGlyPheAlaThrTyrGluLeuLysArgSerTyrAsnTrp 540

Db       1561 TCTTTACAGAATGCAGATGGAGGCCTCGCAACATATGAGCTCACCAGGTCATACCCTTGG 1620

Qy        541 LeuGluIleIleAsnProAlaGluThrPheGlyAspIleValIleAspTyrProTyrVal 560

Db       1621 TTAGAGCTAATCAACCCAGCAGAAACCTTTGGCGATATTGTTATTGATTATCCTTACGTG 1680

Qy        561 GluCysThrSerAlaAlaIleGlnAlaLeuAlaSerPheArgLysLeuTyrProGlyHis 580

Db       1681 GAATGTACATCAGCTGCTATCCAAGCTTTGATATCATTTCGAAAGCTGTATCCTGGTCAT 1740

Qy        581 ArgArgGluGluIleGlnHisCysIleAspLysAlaThrThrPheIleGluLysIleGln 600

Db       1741 CGAAAGAAGGAAGTAGATGAGTGCATTGAGAAGGCGGTTAAGTTCATTGAATCCATTCAA 1800

Qy        601 AlaSerAspGlySerTrpTyrGlySerTrpGlyValCysPheThrTyrGlyAlaTrpPhe 620

Db       1801 GCAGCAGATGGCTCATGGTATGGATCATGGGCTGTTTGCTTCACGTATGGTACGTGGTTT 1860

Qy        621 GlyValLysGlyLeuIleAlaAlaGlyArgSerPheSerAsnCysSerSerIleArgLys 640

Db       1861 GGAGTGAAAGGGCTGGTAGCTGTTGGAAAAACATTGAAAAACTCTCCACATGTTGCTAAA 1920

Qy        641 AlaCysGluPheLeuLeuSerLysGlnLeuProSerGlyGlyTrpGlyGluSerTyrLeu 660

Db       1921 GCTTGTGAATTTCTATTGTCGAAACAACAACCTTCGGGCGGCTGGGGAGAAAGCTATCTT 1980

Qy        661 SerCysGlnAsnLysValTyrSerAsnLeuGluGlyAsnArgSerHisValValAsnThr 680

Db       1981 TCATGTCAAGACAAGGTCTATTCAAACCTTGATGGCAACAGATCTCACGTCGTGAATACA 2040

Qy        681 GlyTrpAlaMetLeuAlaLeuIleAspAlaGlyGlnAlaLysArgAspSerGlnProLeu 700

Db       2041 GCATGGGCTATGCTCGCACTCATTGGTGCTGGGCAAGCTGAGGTAGACCGGAAACCACTA 2100

Qy        701 HisArgAlaAlaAlaTyrLeuIleAsnSerGlnLeuGluAspGlyAspPheProGlnGln 720

Db       2101 CACCGGGCTGCAAGATACTTGATTAATGCTCAAATGGAGAATGGTGATTTTCCACAACAG 2160

Qy        721 GluIleMetGlyValPheAsnLysAsnCysMetIleThrTyrAlaAlaTyrArgAsnIle 740

Db       2161 GAAATAATGGGAGTCTTCAATAGGAACTGCATGATAACATATGCCGCGTATCGAAACATT 2220

Qy        741 PheProIleTrpAlaLeuGlyGluTyrGlnSerGlnValLeu 754

Db       2221 TTTCCGATATGGGCTTTGGGGGAGTACCGTTGTCAGGTATTA 2262
```

RESULT 4
ADE28072
ID ADE28072 standard; cDNA; 2558 BP.
XX
AC ADE28072;
XX
DT 29-JAN-2004 (first entry)
XX
DE Corn cycloartenol synthase cDNA.
XX
KW squalene metabolism; transgenic; plant; genetic; physical gene mapping;
KW breeding; corn; cycloartenol synthase; ss; gene; maize.
XX
OS Zea mays.
XX

FH	Key	Location/Qualifiers
FT	CDS	3. .2129
FT		/*tag= a
FT		/partial
FT		/product= "Corn cycloartenol synthase protein"
FT		/note= "No start codon"

XX

PN US6630617-B1.
XX
PD 07-OCT-2003.
XX
PF 15-OCT-1999; 99US-00419679.
XX
PR 23-OCT-1998; 98US-0105405P.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Famodu OO, Kinney AJ, Rafalski JA;
XX
DR WPI; 2003-810554/76.
DR P-PSDB; ADE28073.
XX
PT New isolated polynucleotide encoding a squalene metabolic enzyme, useful
PT as probes in gene mapping, in producing transgenic plants or in plant
PT breeding to develop lines having desired phenotypes.
XX
PS Disclosure; SEQ ID NO 1; 31pp; English.
XX
CC The invention relates to a novel isolated polynucleotide encoding a
CC squalene metabolic enzyme. The polynucleotide of the invention may be
CC used to create transgenic plants in which the polypeptides are present at
CC higher or lower levels than normal or are found within cell types or
CC developmental stages in which they are not normally found. Furthermore,
CC the polynucleotide may be used as a probe for genetically and physically
CC mapping genes and as a marker for traits linked to those genes, as well
CC as in plant breeding in order to develop lines with desired phenotypes.
CC The current sequence is that of the corn cycloartenol synthase cDNA of
CC the invention.
XX
SQ Sequence 2558 BP; 706 A; 532 C; 636 G; 684 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	6.62e-298	Length:	2558
Score:	3125.00	Matches:	545
Percent Similarity:	88.87%	Conservative:	78
Best Local Similarity:	77.75%	Mismatches:	78
Query Match:	75.34%	Indels:	0
DB:	10	Gaps:	0

US-09-419-679-6 (1-757) x ADE28072 (1-2558)

```
Qy         54 ThrHisLysHisSerAlaAspLeuLeuMetArgMetGlnPheAlaArgGluAsnProThr 73

Db          3 ACGAGCAAGCACAGCGCCGACCTCCTCATGCGCATCCAGTTCGCCAAAGAAAACTCGATT 62

Qy         74 ArgGluValLeuProLysValGlyValLysAspIleGluAspValThrGlnGluIleVal 93

Db         63 GAGCTTCACCTTCCAGGCATCAAGCTCGGTGAGCATGAAGATGTGACCGAGGAAGCTGTG 122

Qy         94 ThrLysThrLeuArgArgAlaValSerPheHisSerThrLeuGlnCysHisAspGlyHis 113

Db        123 TTGACTACATTGAAAAGGGCAATCAGCCGTTTCTCTACTCTCCAGGCACATGATGGACAC 182

Qy        114 TrpProGlyAspTyrGlyGlyProMetPheLeuMetProGlyLeuValIleThrLeuSer 133

Db        183 TGGCCTGGGGATTATGGTGGTCCTATGTTCCTTATGCCAGGCTTGATCATAACATTGTAT 242

Qy        134 IleThrGlyAlaLeuAsnThrValLeuThrGluGluHisArgLysGluIleCysArgTyr 153

Db        243 GTGACTGGAGCACTAAACACTGTCTTGTCATTGGAACACCAGAAGGAGATCCGCCGGTAT 302

Qy        154 LeuTyrAsnHisGlnAsnLysAspGlyGlyTrpGlyLeuHisIleGluGlyProSerThr 173

Db        303 CTTTATAATCACCAGAATGAAGATGGCGGCTGGGGCTTGCACATTGAGGGTCCAAGCACC 362

Qy        174 MetPheGlySerValLeuSerTyrIleThrLeuArgLeuLeuGlyGluGlyProAsnAsp 193

Db        363 ATGTTCGGCTCAGCACTGACCTATGTTATTTTGAGATTGCTTGGAGAGGGACCAGATAGT 422

Qy        194 GlyGlnGlyGluMetGluLysAlaArgAspTrpIleLeuGlyHisGlyGlyAlaThrTyr 213
```

PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US026685.
XX
PR 24-AUG-2000; 2000US-0227866P.
PR 26-JAN-2001; 2001US-0264647P.
PR 22-JUN-2001; 2001US-0300111P.
XX
PA (SCRI) SCRIPPS RES INST.
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Harper JF, Kreps J, Wang X, Zhu T;
XX
DR WPI; 2002-304127/34.
XX
PT Identifying a stress condition to which a plant cell has been exposed and
PT producing plants with increased tolerance to these abiotic stresses.
XX
PS Claim 144; SEQ ID NO 833; 577pp + Sequence Listing; English.
XX
CC The invention relates to identifying a stress condition to which a plant
CC cell has been exposed, comprising: (a) contacting nucleic acid
CC representative of expressed polynucleotides in the plant cell with an
CC array or probes representative of the plant cell genome; and (b)
CC detecting a profile of expressed polynucleotides in the plant cell
CC characteristic of a stress response. The method is useful in the
CC production of transgenic plants, cells and seeds and in producing plants
CC with increased tolerance to abiotic stress. The present sequence is that
CC of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used
CC in methods of the invention. Note: The sequence data for this patent is
CC not represented in the printed specification but is based on sequence
CC information supplied to Derwent by the European Patent Office
XX
SQ Sequence 2280 BP; 653 A; 449 C; 562 G; 616 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	0	Length:	2280
Score:	3480.00	Matches:	607
Percent Similarity:	90.72%	Conservative:	77
Best Local Similarity:	80.50%	Mismatches:	70
Query Match:	83.90%	Indels:	0
DB:	6	Gaps:	0

US-09-419-679-6 (1-757) x ABZ13028 (1-2280)

```
Qy          1 MetTrpLysLeuLysPheAlaGluGlyGlyAsnProTrpLeuArgThrLeuAsnAsnHis 20
              ||||||||||||||   |||||||||||:::|||||||||||||||   |||||||||
Db          1 ATGTGGAAACTGAAGATCGCGGAAGGAGGTAGTCCATGGCTTAGAACCACCAATAATCAC 60

Qy         21 ValGlyArgGlnValTrpGluPheAspProLysLeuGlySerProGlnAspLeuLeuGlu 40
              |||||||||||||   |||||||||||||||   ||||||:::|||:::||||||
Db         61 GTCGGAAGACAGTTTTGGGAGTTCGATCCGAATCTCGGTACTCCTGAGGATCTCGCCGCC 120

Qy         41 IleGluLysAlaArgGlnAsnPheHisAspAsnArgPheThrHisLysHisSerAlaAsp 60
              :::|||:::||||||::::::|||   ||||||||||||      ||||||||||||||||
Db        121 GTCGAAGAAGCTAGGAAGTCTTTTTCGGATAATCGATTCGTGCAGAAACATAGCGCCGAT 180

Qy         61 LeuLeuMetArgMetGlnPheAlaArgGluAsnProThrArgGluValLeuProLysVal 80
              ||||||||||||:::||||||:::|||||||||              |||||||||:::|||
Db        181 CTGCTTATGCGCCTTCAGTTTTCAAGAGAAAATTTGATTAGCCCAGTTTTACCTCAAGTC 240

Qy         81 GlyValLysAspIleGluAspValThrGlnGluIleValThrLysThrLeuArgArgAla 100
              :::::::|||   :::|||||||||:::|||:::|||      ||||||:::|||
Db        241 AAAATCGAAGACACTGATGATGTTACAGAGGAGATGGTGGAAACCACGTTAAAGAGGGGT 300

Qy        101 ValSerPheHisSerThrLeuGlnCysHisAspGlyHisTrpProGlyAspTyrGlyGly 120
              :::   |||:::||||||:::|||   ||||||||||||||||||||||||||||||||
Db        301 CTAGATTTCTATTCAACTATACAGGCACACGACGGGCACTGGCCAGGTGATTATGGTGGT 360

Qy        121 ProMetPheLeuMetProGlyLeuValIleThrLeuSerIleThrGlyAlaLeuAsnThr 140
              ||||||||||||:::||||||||||:::||||||||||||||||||||||||||||||||
Db        361 CCTATGTTTCTTCTCCCAGGACTGATAATTACACTCTCCATAACTGGAGCACTGAATACA 420

Qy        141 ValLeuThrGluGluHisArgLysGluIleCysArgTyrLeuTyrAsnHisGlnAsnLys 160
              ||||||:::|||:::|||:::::|||:::  ||||||||||||||||||||||||:::
Db        421 GTATTGTCGGAACAACATAAACAAGAAATGCGCCGTTATCTCTATAATCACCAGAATGAG 480

Qy        161 AspGlyGlyTrpGlyLeuHisIleGluGlyProSerThrMetPheGlySerValLeuSer 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        481 GACGGAGGTTGGGGTTTACATATTGAGGGCCCTAGCACCATGTTTGGGTCTGTGTTGAAC 540

Qy        181 TyrIleThrLeuArgLeuLeuGlyGluGlyProAsnAspGlyGlnGlyGluMetGluLys 200
              |||:::||||||||||||||||||||||||||||||||||   |||:::|||||||||
Db        541 TATGTTACTCTAAGGTTGCTTGGAGAAGGACCTAACGATGGAGATGGAGATATGGAGAAA 600

Qy        201 AlaArgAspTrpIleLeuGlyHisGlyGlyAlaThrTyrIleThrSerTrpGlyLysMet 220
              |||||||||||||||   ||||||||||||||||   ||||||||||||||||||||
Db        601 GGACGAGACTGGATACTAAATCATGGTGGTGCTACCAATATTACATCTTGGGGGAAAATG 660

Qy        221 TrpLeuSerValLeuGlyValTyrGluTrpSerGlyAsnAsnProLeuProProGluIle 240
              ||||||||||||||||||   :::||||||||||||||||||||||||||||||||||||
Db        661 TGGCTATCGGTACTTGGAGCTTTTGAATGGTCCGGAAATAACCCACTGCCACCTGAGATA 720

Qy        241 TrpLeuLeuProTyrMetLeuProPheHisProGlyArgMetTrpCysHisCysArgMet 260
              ||||||||||||||   ||||||   ||||||||||||||||||||||||||||||||
Db        721 TGGCTTCTCCCATATTTCCTGCCAATACATCCAGGAAGGATGTGGTGCCATTGTCGAATG 780

Qy        261 ValTyrLeuProMetSerTyrLeuTyrGlyLysArgPheValGlyProIleSerProThr 280
              ||||||||||||||||||||||||||||||||||||||||||:::   |||
Db        781 GTGTACTTGCCGATGTCGTATTTGTATGGAAAAAGGTTTGTGGGTCCCATAACGTCCACT 840

Qy        281 ValLeuSerLeuArgLysGluLeuTyrThrValProTyrHisAspIleAspTrpAspGln 300
              ||||||||||||||||||||||:::|||||||||||||:::::::::|||::::::
Db        841 GTTTTATCACTGAGAAAGGAGCTTTTCACAGTACCATATCATGAAGTCAACTGGAATGAA 900

Qy        301 AlaArgAsnLeuCysAlaLysGluAspLeuTyrTyrProHisProLeuValGlnAspIle 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GCACGCAACCTTTGCGCAAAGGAGGATTTATACTACCCACATCCACTTGTGCAAGATATT 960

Qy        321 LeuTrpAlaSerLeuHisLysPheLeuGluProIleLeuMetHisTrpProGlyLysArg 340
              |||||||||||||||||||   :::|||||||:::||||||   |||||||||
Db        961 CTTTGGGCATCACTTCATAAGATTGTTGAGCCTGTTCTGATGCGATGGCCTGGTGCAAAT 1020

Qy        341 LeuArgGluLysAlaIleIleSerAlaLeuGluHisIleHisTyrGluAspGluAsnThr 360
              ||||||||||||||||||   :::|||:::||||||||||||||||||||||||||||
Db       1021 TTGAGAGAAAAGGCTATAAGAACCGCAATAGAACATATTCATTATGAAGATGAGAATACT 1080

Qy        361 ArgTyrIleCysIleGlyProValAsnLysValLeuAsnMetLeuCysCysTrpValGlu 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AGGTACATCTGCATAGGTCCCGTGAACAAGGTATTAAATATGCTTTGCTGTTGGGTAGAA 1140

Qy        381 AspProAsnSerGluAlaPheLysLeuHisLeuProArgIleTyrAspTyrLeuTrpIle 400
              ||||||||||||||||||||||||||||||||||||||||||:::|||:::|||||:::
Db       1141 GACCCAAACTCAGAGGCTTTCAAGTTGCACCTACCAAGAATCCATGACTTTCTCTGGTTA 1200

Qy        401 AlaGluAspGlyMetLysMetGlnGlyTyrAsnGlySerGlnLeuTrpAspThrAlaPhe 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db       1201 GCTGAAGATGGAATGAAGATGCAGGGTTATAACGGAAGCCAGCTATGGGATACAGGTTTT 1260

Qy        421 AlaValGlnAlaIleIleAlaSerAsnLeuIleGluGluPheGlyProThrIleArgLys 440
              |||:::|||||||||:::|||:::|||||||:::|||||||:::|||||   :::   |||
Db       1261 GCTATTCAAGCGATTTTGGCAACTAACCTCGTCGAAGAATATGGGCCCGTTTTGGAAAAA 1320

Qy        441 AlaHisThrTyrIleLysAsnSerGlnValLeuGluAspCysProGlyAspLeuAsnLys 460
              ||||||:::::::::|||||||||||||||||||||||||||||||||||||||||||
Db       1321 GCACATTCATTTGTCAAGAATTCCCAGGTGTTAGAAGACTGCCCTGGAGATCTGAATTAC 1380

Qy        461 TrpTyrArgHisIleSerLysGlyAlaTrpProPheSerThrGlyAspHisGlyTrpPro 480
              ||||||||||||||||||||||||||||||||||   ||||||||||||||||
Db       1381 TGGTATCGCCACATTTCTAAAGGGGCTTGGCCTTTCTCAACTGCAGATCACGGTTGGCCC 1440

Qy        481 IleSerAspCysThrAlaGluGlyLeuLysAlaValLeuLeuLeuSerLysIleAlaPro 500
              ||||||||||||||||||||||||||||||||   ||||||||||||||:::
Db       1441 ATCTCTGACTGCACCGCAGAAGGACTGAAAGCTGCTCTTTTGCTATCCAAAGTTCCCAAG 1500

Qy        501 GluIleValGlyGluProIleAspValLysArgLeuTyrAspSerValAsnValIleLeu 520
```

```
Db        325 GTAAGTTTCCATTCAACTCTCCAGTGCCATGACGGACACTGGCCGGGAGATTATGGAGGT 384

Qy        121 ProMetPheLeuMetProGlyLeuValIleThrLeuSerIleThrGlyAlaLeuAsnThr 140

Db        385 CCCATGTTTCTGATGCCTGGCTTGGTAATTACTCTGTCTATCACTGGGGCGTTGAATACA 444

Qy        141 ValLeuThrGluGluHisArgLysGluIleCysArgTyrLeuTyrAsnHisGlnAsnLys 160

Db        445 GTCTTAACTGAAGAACATAGAAAGGAAATATGCCGTTACCTCTATAATCATCAAAACAAG 504

Qy        161 AspGlyGlyTrpGlyLeuHisIleGluGlyProSerThrMetPheGlySerValLeuSer 180

Db        505 GATGGTGGGTGGGGTTTGCATATTGAAGGTCCAAGCACCATGTTTGGCTCTGTCTTGAGT 564

Qy        181 TyrIleThrLeuArgLeuLeuGlyGluGlyProAsnAspGlyGlnGlyGluMetGluLys 200

Db        565 TATATTACTCTGAGATTGCTAGGTGAGGGGCCTAATGATGGACAAGGGGAAATGGAGAAG 624

Qy        201 AlaArgAspTrpIleLeuGlyHisGlyGlyAlaThrTyrIleThrSerTrpGlyLysMet 220

Db        625 GCACGTGACTGGATTCTAGGGCATGGTGGTGCTACTTATATAACGTCATGGGGGAAGATG 684

Qy        221 TrpLeuSerValLeuGlyValTyrGluTrpSerGlyAsnAsnProLeuProProGluIle 240

Db        685 TGGCTTTCAGTACTTGGAGTGTATGAATGGTCTGGAAATAATCCCCTGCCCCCTGAGATA 744

Qy        241 TrpLeuLeuProTyrMetLeuProPheHisProGlyArgMetTrpCysHisCysArgMet 260

Db        745 TGGCTCCTTCCATACATGCTTCCATTTCATCCAGGAAGGATGTGGTGTCACTGCCGGATG 804

Qy        261 ValTyrLeuProMetSerTyrLeuTyrGlyLysArgPheValGlyProIleSerProThr 280

Db        805 GTCTATTTGCCGATGTCCTACTTATATGGCAAGAGGTTTGTTGGTCCAATCTCACCAACA 864

Qy        281 ValLeuSerLeuArgLysGluLeuTyrThrValProTyrHisAspIleAspTrpAspGln 300

Db        865 GTATTATCTTTGAGAAAAGAGCTTTATACAGTACCATACCATGATATAGATTGGGATCAG 924

Qy        301 AlaArgAsnLeuCysAlaLysGluAspLeuTyrTyrProHisProLeuValGlnAspIle 320

Db        925 GCTCGCAATTTGTGTGCAAAGGAAGATTTGTACTATCCTCACCCACTTGTACAGGATATT 984

Qy        321 LeuTrpAlaSerLeuHisLysPheLeuGluProIleLeuMetHisTrpProGlyLysArg 340

Db        985 CTTTGGGCATCTCTACACAAGTTCCTTGAGCCTATTCTGATGCATTGGCCTGGAAAAAGA 1044

Qy        341 LeuArgGluLysAlaIleIleSerAlaLeuGluHisIleHisTyrGluAspGluAsnThr 360

Db       1045 TTGAGGGAAAAGGCTATTATTTCTGCATTGGAGCATATACATTACGAAGATGAGAATACT 1104

Qy        361 ArgTyrIleCysIleGlyProValAsnLysValLeuAsnMetLeuCysCysTrpValGlu 380

Db       1105 CGATATATTTGCATAGGTCCTGTAAATAAGGTGTTAAATATGCTTTGCTGTTGGGTGGAA 1164

Qy        381 AspProAsnSerGluAlaPheLysLeuHisLeuProArgIleTyrAspTyrLeuTrpIle 400

Db       1165 GATCCAAATTCTGAGGCCTTCAAGTTGCATCTTCCCAGGATTTATGATTATCTATGGATT 1224

Qy        401 AlaGluAspGlyMetLysMetGlnGlyTyrAsnGlySerGlnLeuTrpAspThrAlaPhe 420

Db       1225 GCAGAAGATGGCATGAAAATGCAGGGCTACAATGGAAGTCAACTATGGGACACTGCTTTT 1284

Qy        421 AlaValGlnAlaIleIleAlaSerAsnLeuIleGluGluPheGlyProThrIleArgLys 440

Db       1285 GCTGTCCAAGCAATTATTGCATCTAACCTCATTGAAGAATTTGGTCCAACTATAAGAAAA 1344

Qy        441 AlaHisThrTyrIleLysAsnSerGlnValLeuGluAspCysProGlyAspLeuAsnLys 460

Db       1345 GCTCATACCTATATTAAGAATTCACAGGTTTTAGAAGATTGTCCAGGTGATCTTAATAAA 1404

Qy        461 TrpTyrArgHisIleSerLysGlyAlaTrpProPheSerThrGlyAspHisGlyTrpPro 480

Db       1405 TGGTACCGTCACATTTCAAAAGGTGCTTGGCCTTTTTCAACTGGAGATCATGGATGGCCA 1464

Qy        481 IleSerAspCysThrAlaGluGlyLeuLysAlaValLeuLeuLeuSerLysIleAlaPro 500

Db       1465 ATTTCTGACTGCACAGCTGAAGGACTGAAAGCTGTTCTATTACTATCCAAAATTGCACCA 1524

Qy        501 GluIleValGlyGluProIleAspValLysArgLeuTyrAspSerValAsnValIleLeu 520

Db       1525 GAAATAGTTGGTGAGCCAATAGACGTGAAGCGATTATATGATTCTGTAAATGTCATTCTC 1584

Qy        521 SerLeuGlnAsnGluAspGlyGlyPheAlaThrTyrGluLeuLysArgSerTyrAsnTrp 540

Db       1585 TCACTACAGAATGAAGATGGTGGTTTTGCAACATATGAGCTTAAACGATCTTATAATTGG 1644

Qy        541 LeuGluIleIleAsnProAlaGluThrPheGlyAspIleValIleAspTyrProTyrVal 560

Db       1645 TTGGAGATAATCAATCCTGCTGAAACTTTGGTGACATCGTTATTGATTATCCTTATGTG 1704

Qy        561 GluCysThrSerAlaAlaIleGlnAlaLeuAlaSerPheArgLysLeuTyrProGlyHis 580

Db       1705 GAATGTACATCAGCAGCGATTCAAGCTTTGGCATCATTTAGGAAATTATATCCTGGGCAT 1764

Qy        581 ArgArgGluGluIleGlnHisCysIleAspLysAlaThrThrPheIleGluLysIleGln 600

Db       1765 CGCCGAGAAGAAATACAACATTGTATCGATAAAGCCACTACCTTCATTGAAAAAATACAA 1824

Qy        601 AlaSerAspGlySerTrpTyrGlySerTrpGlyValCysPheThrTyrGlyAlaTrpPhe 620

Db       1825 GCTTCAGATGGATCATGGTATGGTTCTTGGGGAGTTTGCTTCACTTACGGTGCTTGGTTT 1884

Qy        621 GlyValLysGlyLeuIleAlaAlaGlyArgSerPheSerAsnCysSerSerIleArgLys 640

Db       1885 GGGGTAAAAGGTCTGATTGCTGCTGGAAGGAGTTTCAGTAATTGCTCAAGCATCCGTAAA 1944

Qy        641 AlaCysGluPheLeuLeuSerLysGlnLeuProSerGlyGlyTrpGlyGluSerTyrLeu 660

Db       1945 GCTTGTGAATTTCTGCTGTCCAAGCAGCTTCCTTCTGGTGGCTGGGGAGAGAGTTATCTG 2004

Qy        661 SerCysGlnAsnLysValTyrSerAsnLeuGluGlyAsnArgSerHisValValAsnThr 680

Db       2005 TCCTGTCAAAACAAGGTGTATTCAAATCTGGAAGGCAACAGGTCTCATGTGGTCAACACT 2064

Qy        681 GlyTrpAlaMetLeuAlaLeuIleAspAlaGlyGlnAlaLysArgAspSerGlnProLeu 700

Db       2065 GGGTGGGCTATGTTGGCTCTCATTGATGCTGGACAGGCTAAGAGAGATTCGCAACCACTG 2124

Qy        701 HisArgAlaAlaAlaTyrLeuIleAsnSerGlnLeuGluAspGlyAspProGlnGln 720

Db       2125 CACCGGGCAGCTGCATACTTGATAAATTCCCAATTGGAGGACGGTGACTTTCCGCAGCAG 2184

Qy        721 GluIleMetGlyValPheAsnLysAsnCysMetIleThrTyrAlaAlaTyrArgAsnIle 740

Db       2185 GAAATAATGGGAGTCTTCAACAAGAATTGCATGATCACATACGCCGCATACAGAAACATA 2244

Qy        741 ProIleTrpAlaLeuGlyGluTyrGlnSerGlnValLeuGlnSerArg 757

Db       2245 TTCCCCATTTGGGCGTTGGGAGAATACCAATCCCAAGTATTGCAATCTCGT 2295
```

RESULT 3
ABZ13028
ID ABZ13028 standard; DNA; 2280 BP.
XX
AC ABZ13028;
XX
DT 21-JAN-2003 (first entry)
XX
DE Arabidopsis thaliana stress regulated gene SEQ ID NO 833.
XX
KW Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
XX
OS Arabidopsis thaliana.
XX
PN WO200216655-A2.
XX

```
Qy        561 GluCysThrSerAlaAlaIleGlnAlaLeuAlaSerPheArgLysLeuTyrProGlyHis 580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1705 GAATGTACATCAGCAGCGATTCAAGCTTTGGCATCATTTAGGAAATTATATCCTGGGCAT 1764

Qy        581 ArgArgGluGluIleGlnHisCysIleAspLysAlaThrThrPheIleGluLysIleGln 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1765 CGCCGAGAAGAAATACAACATTGTATCGATAAAGCCACTACCTTCATTGAAAAAATACAA 1824

Qy        601 AlaSerAspGlySerTrpTyrGlySerTrpGlyValCysPheThrTyrGlyAlaTrpPhe 620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1825 GCTTCAGATGGATCATGGTATGGTTCTTGGGGAGTTTGCTTCACTTACGGTGCTTGGTTT 1884

Qy        621 GlyValLysGlyLeuIleAlaAlaGlyArgSerPheSerAsnCysSerSerIleArgLys 640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1885 GGGGTAAAAGGTCTGATTGCTGCTGGAAGGAGTTTCAGTAATTGCTCAAGCATCCGTAAA 1944

Qy        641 AlaCysGluPheLeuLeuSerLysGlnLeuProSerGlyGlyTrpGlyGluSerTyrLeu 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1945 GCTTGTGAATTTCTGCTGTCCAAGCAGCTTCCTTCTGGTGGCTGGGGAGAGAGTTATCTG 2004

Qy        661 SerCysGlnAsnLysValTyrSerAsnLeuGluGlyAsnArgSerHisValValAsnThr 680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2005 TCCTGTCAAAACAAGGTGTATTCAAATCTGGAAGGCAACAGGTCTCATGTGGTCAACACT 2064

Qy        681 GlyTrpAlaMetLeuAlaLeuIleAspAlaGlyGlnAlaLysArgAspSerGlnProLeu 700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2065 GGGTGGGCTATGTTGGCTCTCATTGATGCTGGACAGGCTAAGAGAGATTCGCAACCACTG 2124

Qy        701 HisArgAlaAlaAlaTyrLeuIleAsnSerGlnLeuGluAspGlyAspPheProGlnGln 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2125 CACCGGGCAGCTGCATACTTGATAAATTCCCAATTGGAGGACGGTGACTTTCCGCAGCAG 2184

Qy        721 GluIleMetGlyValPheAsnLysAsnCysMetIleThrTyrAlaAlaTyrArgAsnIle 740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2185 GAAATAATGGGAGTCTTCAACAAGAATTGCATGATCACATACGCCGCATACAGAAACATA 2244

Qy        741 ProIleTrpAlaLeuGlyGluTyrGlnSerGlnValLeuGlnSerArg 757
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db       2245 TTCCCCATTTGGGCGTTGGGAGAATACCAATCCCAAGTATTGCAATCTCGT 2295
```

RESULT 2
ADQ95187
ID ADQ95187 standard; cDNA; 2560 BP.
XX
AC ADQ95187;
XX
DT 23-SEP-2004 (first entry)
XX
DE Soybean cycloartenol synthase cDNA.
XX
KW soybean; cycloartenol synthase; squalene metabolism; gene; ss.
XX
OS Glycine max.
XX
FH Key Location/Qualifiers
FT CDS 25. .2298
FT /*tag= a
FT /product= "Soybean cycloartenol synthase"
XX
PN US2004121439-A1.
XX
PD 24-JUN-2004.
XX
PF 27-JUN-2003; 2003US-00607726.
XX
PR 23-OCT-1998; 98US-0105405P.
PR 15-OCT-1999; 99US-00419679.
XX
PA (FALC/) FALCO S C.
PA (FAMO/) FAMODU O O.
PA (HITZ/) HITZ W D.
PA (KINN/) KINNEY A J.
PA (RAFA/) RAFALSKI J A.
PA (MCGO/) MCGONIGLE B.
PA (LOHM/) LOHMAN K N.
XX
PI Falco SC, Famodu OO, Hitz WD, Kinney AJ, Rafalski JA;
PI Mcgonigle B, Lohman KN;
XX
DR WPI; 2004-479684/45.
DR P-PSDB; ADQ95188.
XX
PT New composition comprising enzymes involved in squalene metabolism,
PT useful in altering squalene metabolism in plants and seeds.
XX
PS Claim 2; SEQ ID NO 5; 33pp; English.
XX
CC The invention describes a composition comprising an isolated
CC polynucleotide comprising a nucleotide sequence encoding a first
CC polypeptide of at least 200 amino acids. A composition comprising an
CC isolated polynucleotide comprising a nucleotide sequence encoding a first
CC polypeptide of at least 200 amino acids that has at least 90% identity
CC based on the Clustal method of alignment when compared to a polypeptide
CC comprising a sequence of 701, 541, 757, 345, 386, 330, 523, or 467 amino
CC acids (EVEN SEQ ID NOS: 2-16), or an isolated polynucleotide comprising
CC the complement of the nucleotide sequence, or comprising a polypeptide of
CC at least 200 amino acids that has at least 90% identity based on the
CC Clustal method of alignment when compared to a polypeptide of EVEN SEQ ID
CC NOS: 2-16. Also described are: a chimeric gene comprising the isolated
CC polynucleotide described above operably linked to suitable regulatory
CC sequences; an isolated host cell comprising the chimeric gene of (1) or
CC the isolated polynucleotide described above; a virus comprising the
CC isolated polynucleotide described above; selecting an isolated
CC polynucleotide that affects the level of expression of a polypeptide in a
CC host or plant cell; obtaining a nucleic acid fragment encoding a
CC polypeptide; and evaluating at least one compound for its ability to
CC inhibit the activity of a polypeptide. The composition, methods,
CC polynucleotides and polypeptides are useful in altering squalene
CC metabolism in plants and seeds. This sequence encodes soybean
CC cycloartenol synthase, a member of the squalene metabolism pathway.
XX
SQ Sequence 2560 BP; 747 A; 483 C; 596 G; 734 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	0	Length:	2560
Score:	4148.00	Matches:	757
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	12	Gaps:	0

US-09-419-679-6 (1-757) x ADQ95187 (1-2560)

```
Qy          1 MetTrpLysLeuLysPheAlaGluGlyGlyAsnProTrpLeuArgThrLeuAsnAsnHis 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         25 ATGTGGAAGCTCAAGTTCGCCGAAGGAGGGAATCCATGGCTTCGGACATTGAACAATCAC 84

Qy         21 ValGlyArgGlnValTrpGluPheAspProLysLeuGlySerProGlnAspLeuLeuGlu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         85 GTTGGAAGACAGGTGTGGGAGTTCGATCCTAAGCTTGGATCGCCGCAAGATCTCCTCGAG 144

Qy         41 IleGluLysAlaArgGlnAsnPheHisAspAsnArgPheThrHisLysHisSerAlaAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        145 ATTGAGAAAGCTCGCCAGAATTTTCACGATAACCGCTTTACCCACAAACACAGCGCTGAT 204

Qy         61 LeuLeuMetArgMetGlnPheAlaArgGluAsnProThrArgGluValLeuProLysVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        205 CTACTTATGCGGATGCAGTTCGCAAGAGAGAACCCAACACGTGAAGTCTTGCCCAAAGTC 264

Qy         81 GlyValLysAspIleGluAspValThrGlnGluIleValThrLysThrLeuArgArgAla 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        265 GGAGTTAAGGATATTGAGGATGTGACCCAAGAGATTGTGACAAAAACATTAAGAAGGGCC 324

Qy        101 ValSerPheHisSerThrLeuGlnCysHisAspGlyHisTrpProGlyAspTyrGlyGly 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

XX
PI Famodu OO, Kinney AJ, Rafalski JA;
XX
DR WPI; 2003-810554/76.
DR P-PSDB; ADE28077.
XX
PT New isolated polynucleotide encoding a squalene metabolic enzyme, useful
PT as probes in gene mapping, in producing transgenic plants or in plant
PT breeding to develop lines having desired phenotypes.
XX
PS Disclosure; SEQ ID NO 5; 31pp; English.
XX
CC The invention relates to a novel isolated polynucleotide encoding a
CC squalene metabolic enzyme. The polynucleotide of the invention may be
CC used to create transgenic plants in which the polypeptides are present at
CC higher or lower levels than normal or are found within cell types or
CC developmental stages in which they are not normally found. Furthermore,
CC the polynucleotide may be used as a probe for genetically and physically
CC mapping genes and as a marker for traits linked to those genes, as well
CC as in plant breeding in order to develop lines with desired phenotypes.
CC The current sequence is that of the soybean cycloartenol synthase cDNA of
CC the invention.
XX
SQ Sequence 2560 BP; 747 A; 483 C; 596 G; 734 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	0	Length:	2560
Score:	4148.00	Matches:	757
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	10	Gaps:	0

US-09-419-679-6 (1-757) x ADE28076 (1-2560)

```
Qy          1 MetTrpLysLeuLysPheAlaGluGlyGlyAsnProTrpLeuArgThrLeuAsnAsnHis 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         25 ATGTGGAAGCTCAAGTTCGCCGAAGGAGGGAATCCATGGCTTCGGACATTGAACAATCAC 84

Qy         21 ValGlyArgGlnValTrpGluPheAspProLysLeuGlySerProGlnAspLeuLeuGlu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         85 GTTGGAAGACAGGTGTGGGAGTTCGATCCTAAGCTTGGATCGCCGCAAGATCTCCTCGAG 144

Qy         41 IleGluLysAlaArgGlnAsnPheHisAspAsnArgPheThrHisLysHisSerAlaAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        145 ATTGAGAAAGCTCGCCAGAATTTTCACGATAACCGCTTTACCCACAAACACAGCGCTGAT 204

Qy         61 LeuLeuMetArgMetGlnPheAlaArgGluAsnProThrArgGluValLeuProLysVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        205 CTACTTATGCGGATGCAGTTCGCAAGAGAGAACCCAACACGTGAAGTCTTGCCCAAAGTC 264

Qy         81 GlyValLysAspIleGluAspValThrGlnGluIleValThrLysThrLeuArgArgAla 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        265 GGAGTTAAGGATATTGAGGATGTGACCCAAGAGATTGTGACAAAAACATTAAGAAGGGCC 324

Qy        101 ValSerPheHisSerThrLeuGlnCysHisAspGlyHisTrpProGlyAspTyrGlyGly 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        325 GTAAGTTTCCATTCAACTCTCCAGTGCCATGACGGACACTGGCCGGGAGATTATGGAGGT 384

Qy        121 ProMetPheLeuMetProGlyLeuValIleThrLeuSerIleThrGlyAlaLeuAsnThr 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        385 CCCATGTTTCTGATGCCTGGCTTGGTAATTACTCTGTCTATCACTGGGGCGTTGAATACA 444

Qy        141 ValLeuThrGluGluHisArgLysGluIleCysArgTyrLeuTyrAsnHisGlnAsnLys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        445 GTCTTAACTGAAGAACATAGAAAGGAAATATGCCGTTACCTCTATAATCATCAAAACAAG 504

Qy        161 AspGlyGlyTrpGlyLeuHisIleGluGlyProSerThrMetPheGlySerValLeuSer 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        505 GATGGTGGGTGGGGTTTGCATATTGAAGGTCCAAGCACCATGTTTGGCTCTGTCTTGAGT 564

Qy        181 TyrIleThrLeuArgLeuLeuGlyGluGlyProAsnAspGlyGlnGlyGluMetGluLys 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        565 TATATTACTCTGAGATTGCTAGGTGAGGGGCCTAATGATGGACAAGGGGAAATGGAGAAG 624

Qy        201 AlaArgAspTrpIleLeuGlyHisGlyGlyAlaThrTyrIleThrSerTrpGlyLysMet 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        625 GCACGTGACTGGATTCTAGGGCATGGTGGTGCTACTTATATAACGTCATGGGGGAAGATG 684

Qy        221 TrpLeuSerValLeuGlyValTyrGluTrpSerGlyAsnAsnProLeuProProGluIle 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        685 TGGCTTTCAGTACTTGGAGTGTATGAATGGTCTGGAAATAATCCCCTGCCCCCTGAGATA 744

Qy        241 TrpLeuLeuProTyrMetLeuProPheHisProGlyArgMetTrpCysHisCysArgMet 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        745 TGGCTCCTTCCATACATGCTTCCATTTCATCCAGGAAGGATGTGGTGTCACTGCCGGATG 804

Qy        261 ValTyrLeuProMetSerTyrLeuTyrGlyLysArgPheValGlyProIleSerProThr 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        805 GTCTATTTGCCGATGTCCTACTTATATGGCAAGAGGTTTGTTGGTCCAATCTCACCAACA 864

Qy        281 ValLeuSerLeuArgLysGluLeuTyrThrValProTyrHisAspIleAspTrpAspGln 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        865 GTATTATCTTTGAGAAAAGAGCTTTATACAGTACCATACCATGATATAGATTGGGATCAG 924

Qy        301 AlaArgAsnLeuCysAlaLysGluAspLeuTyrTyrProHisProLeuValGlnAspIle 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        925 GCTCGCAATTTGTGTGCAAAGGAAGATTTGTACTATCCTCACCCACTTGTACAGGATATT 984

Qy        321 LeuTrpAlaSerLeuHisLysPheLeuGluProIleLeuMetHisTrpProGlyLysArg 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        985 CTTTGGGCATCTCTACACAAGTTCCTTGAGCCTATTCTGATGCATTGGCCTGGAAAAAGA 1044

Qy        341 LeuArgGluLysAlaIleIleSerAlaLeuGluHisIleHisTyrGluAspGluAsnThr 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1045 TTGAGGGAAAAGGCTATTATTTCTGCATTGGAGCATATACATTACGAAGATGAGAATACT 1104

Qy        361 ArgTyrIleCysIleGlyProValAsnLysValLeuAsnMetLeuCysCysTrpValGlu 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1105 CGATATATTTGCATAGGTCCTGTAAATAAGGTGTTAAATATGCTTTGCTGTTGGGTGGAA 1164

Qy        381 AspProAsnSerGluAlaPheLysLeuHisLeuProArgIleTyrAspTyrLeuTrpIle 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1165 GATCCAAATTCTGAGGCCTTCAAGTTGCATCTTCCCAGGATTTATGATTATCTATGGATT 1224

Qy        401 AlaGluAspGlyMetLysMetGlnGlyTyrAsnGlySerGlnLeuTrpAspThrAlaPhe 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1225 GCAGAAGATGGCATGAAAATGCAGGGCTACAATGGAAGTCAACTATGGGACACTGCTTTT 1284

Qy        421 AlaValGlnAlaIleIleAlaSerAsnLeuIleGluGluPheGlyProThrIleArgLys 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1285 GCTGTCCAAGCAATTATTGCATCTAACCTCATTGAAGAATTTGGTCCAACTATAAGAAAA 1344

Qy        441 AlaHisThrTyrIleLysAsnSerGlnValLeuGluAspCysProGlyAspLeuAsnLys 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1345 GCTCATACCTATATTAAGAATTCACAGGTTTTAGAAGATTGTCCAGGTGATCTTAATAAA 1404

Qy        461 TrpTyrArgHisIleSerLysGlyAlaTrpProPheSerThrGlyAspHisGlyTrpPro 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1405 TGGTACCGTCACATTTCAAAAGGTGCTTGGCCTTTTTCAACTGGAGATCATGGATGGCCA 1464

Qy        481 IleSerAspCysThrAlaGluGlyLeuLysAlaValLeuLeuLeuSerLysIleAlaPro 500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1465 ATTTCTGACTGCACAGCTGAAGGACTGAAAGCTGTTCTATTACTATCCAAAATTGCACCA 1524

Qy        501 GluIleValGlyGluProIleAspValLysArgLeuTyrAspSerValAsnValIleLeu 520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1525 GAAATAGTTGGTGAGCCAATAGACGTGAAGCGATTATATGATTCTGTAAATGTCATTCTC 1584

Qy        521 SerLeuGlnAsnGluAspGlyGlyPheAlaThrTyrGluLeuLysArgSerTyrAsnTrp 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1585 TCACTACAGAATGAAGATGGTGGTTTTGCAACATATGAGCTTAAACGATCTTATAATTGG 1644

Qy        541 LeuGluIleIleAsnProAlaGluThrPheGlyAspIleValIleAspTyrProTyrVal 560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1645 TTGGAGATAATCAATCCTGCTGAAACTTTTGGTGACATCGTTATTGATTATCCTTATGTG 1704
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 5, 2005, 06:54:23 ; Search time 1004 Seconds
(without alignments)
4463.391 Million cell updates/sec

Title: US-09-419-679-6
Perfect score: 4148
Sequence: 1 MWKLKFAEGGNPWLRTLNNH.....RNIFPIWALGEYQSQVLQSR 757

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/6630617/runat_02112005_120300_15067/app_query.fasta_1.903
-DB=N_Geneseq_16Dec04 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=6630617_@CGN_1_1_479_@runat_02112005_120300_15067 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4148	100.0	2560	10	ADE28076	Ade28076 Soybean c
2	4148	100.0	2560	12	ADQ95187	Adq95187 Soybean c
3	3480	83.9	2280	6	ABZ13028	Abz13028 Arabidops
4	3125	75.3	2558	10	ADE28072	Ade28072 Corn cycl
5	3125	75.3	2558	12	ADQ95183	Adq95183 Corn cycl
6	2551	61.5	2538	4	AAD16896	Aad16896 Wheat oxi
7	2538	61.2	2478	4	AAD16894	Aad16894 Soybean b
8	2538	61.2	2478	12	ADF89882	Adf89882 G. max be
9	2516.5	60.7	2093	12	ADI45184	Adi45184 Rice isop
10	2454.5	59.2	2731	4	AAD16893	Aad16893 Soybean o
11	2436	58.7	2766	4	AAD16895	Aad16895 Soybean o
12	2436	58.7	2766	12	ADF89879	Adf89879 G. max ox
13	2433	58.7	1882	10	ADE28074	Ade28074 Rice cycl
14	2433	58.7	1882	12	ADQ95185	Adq95185 Rice cycl
15	2418.5	58.3	2532	3	AAC46759	Aac46759 Arabidops
16	2354.5	56.8	2457	4	AAS09680	Aas09680 Oat cDNA
17	2311.5	55.7	2067	10	ADF89798	Adf89798 M. trunca
18	2300	55.4	2620	12	ADM28673	Adm28673 DNA encod
19	2234	53.9	2493	3	AAC42686	Aac42686 Arabidops
20	2167	52.2	2535	12	ADM11044	Adm11044 Ginseng d
21	2083	50.2	2639	12	ADJ48244	Adj48244 Maize oil
22	1882	45.4	1546	12	ADI45617	Adi45617 Wheat iso
23	1774.5	42.8	1926	12	ADI45430	Adi45430 Rice isop
24	1774.5	42.8	1926	12	ADJ39566	Adj39566 Plant cDN
25	1680.5	40.5	2851	10	ADB53186	Adb53186 Primary r
26	1666.5	40.2	2669	2	AAT59599	Aat59599 Rat-deriv
27	1653.5	39.9	2637	8	ACA89902	Aca89902 Gene diff
28	1653.5	39.9	3206	6	ABK84268	Abk84268 Human cDN
29	1653.5	39.9	3206	10	ADH28966	Adh28966 Human chr
30	1653.5	39.9	3206	12	ADP05441	Adp05441 Human lan
31	1653.5	39.9	3206	13	ACN40868	Acn40868 Tumour-as
32	1653.5	39.9	4346	12	ADE76947	Ade76947 Human cDN
33	1597	38.5	4239	10	ABS56199	Abs56199 cDNA enco
34	1597	38.5	4239	13	ADS16409	Ads16409 Human lan
35	1510	36.4	2160	10	ADB69885	Adb69885 C. neofor
36	1452.5	35.0	2256	8	ABT19094	Abt19094 Aspergill
37	1442	34.8	1300	10	ADE28078	Ade28078 Bread whe
38	1442	34.8	1300	12	ADQ95189	Adq95189 Wheat cyc
39	1425.5	34.4	2208	8	ABT20914	Abt20914 Aspergill
40	1364.5	32.9	2398	8	ABT18500	Abt18500 Aspergill
41	1364.5	32.9	2398	8	ABT20316	Abt20316 Aspergill
42	1364.5	32.9	4398	8	ABT17906	Abt17906 Aspergill
43	1364.5	32.9	4398	8	ABT19720	Abt19720 Aspergill
44	1364.5	32.9	4480	8	ABX63079	Abx63079 Human cDN
45	1303	31.4	2377	13	ADS48639	Ads48639 Bacterial

ALIGNMENTS

RESULT 1
ADE28076
ID ADE28076 standard; cDNA; 2560 BP.
XX
AC ADE28076;
XX
DT 29-JAN-2004 (first entry)
XX
DE Soybean cycloartenol synthase cDNA.
XX
KW squalene metabolism; transgenic; plant; genetic; physical gene mapping;
KW breeding; soybean; cycloartenol synthase; ss; gene.
XX
OS Glycine max.
XX
FH Key Location/Qualifiers
FT CDS 25. .2298
FT /*tag= a
FT /product= "Soybean cycloartenol synthase protein"
XX
PN US6630617-B1.
XX
PD 07-OCT-2003.
XX
PF 15-OCT-1999; 99US-00419679.
XX
PR 23-OCT-1998; 98US-0105405P.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.

/clone="POCBM46"
/tissue_type="callus"
/lab_host="DH10B-TonA"
/clone_lib="potato callus cDNA library, normalized and full-length"
/note="Vector: pCMVSport6.1; Site_1: EcoRI; Site_2: NotI; supplier: RNA was isolated from Solanum tuberosum var. Kennebec callus tissue grown on solid media."

ORIGIN

Query Match 21.7%; Score 556; DB 7; Length 958;
Best Local Similarity 74.5%; Pred. No. 1.1e-138;
Matches 700; Conservative 0; Mismatches 240; Indels 0; Gaps 0;

```
Qy         22 AGAATGTGGAAGCTCAAGTTCGCCGAAGGAGGGAATCCATGGCTTCGGACATTGAACAAT 81
              |||||||||||| | ||  | || |||||||||| |||||||| ||||||  |||||||
Db         19 AGAATGTGGAAGTTGAAAGTAGCAGAAGGAGGGAGTCCATGGCTGCGGACAACGAACAAT 78

Qy         82 CACGTTGGAAGACAGGTGTGGGAGTTCGATCCTAAGCTTGGATCGCCGCAAGATCTCCTC 141
              || || || ||  | || |||||||| || ||  | |||||||| ||  ||||||
Db         79 CATGTAGGTAGGGAAGTATGGGAGTTTGAACCAGACCTTGGATCTCCAGAAGATCGGGCG 138

Qy        142 GAGATTGAGAAAGCTCGCCAGAATTTTCACGATAACCGCTTTACCCACAAACACAGCGCT 201
              ||||| || ||   ||| |  | ||   | |  | || ||     || |||||||| |||
Db        139 GAGATCGAAAAGTTTCGTGAACACTTCACCAAACATCGTTTGGAGCAAAAACACAGTGCT 198

Qy        202 GATCTACTTATGCGGATGCAGTTCGCAAGAGAGAACCCAACACGTGAAGTCTTGCCCAAA 261
              ||||| ||||||||     |||||  |||  |||||||     |  |     |||||||  ||
Db        199 GATCTCCTTATGCGTTACCAGTTGTCAAAGGAGAACCAGGGAATTAGTATCTTGCCACAA 258

Qy        262 GTCGGAGTTAAGGATATTGAGGATGTGACCCAAGAGATTGTGACAAAAACATTAAGAAGG 321
              |||  | |  | || | ||| | | | ||  |||| | |||  | |  ||  | |||||
Db        259 GTCAAAATACATGACAATGAAGTTATTACAGAAGATAGTGTAGCTACCACGCTGAGAAGA 318

Qy        322 GCCGTAAGTTTCCATTCAACTCTCCAGTGCCATGACGGACACTGGCCGGGAGATTATGGA 381
              ||  |||||||  |||| |||||  ||    |||||||| |||||| | ||||||||||||
Db        319 GCGCTAAGTTTTTATTCCACTCTACAAAATCATGACGGTCACTGGGCTGGAGATTATGGA 378

Qy        382 GGTCCCATGTTTCTGATGCCTGGCTTGGTAATTACTCTGTCTATCACTGGGGCGTTGAAT 441
              ||  || ||||||||  |||||||||  ||  |  |||  |||| |||||  ||  |||||   | |||
Db        379 GGCCCGATGTTTCTAATGCCTGGCATGATTATTGCTCTATCTATTACGGGGGCACTCAAT 438

Qy        442 ACAGTCTTAACTGAAGAACATAGAAAGGAAATATGCCGTTACCTCTATAATCATCAAAAC 501
               ||||  ||   |    |||||||  |    ||  ||     ||  ||  ||  |||||  |||||  |||
Db        439 GCAGTATTCTCAAGTGAACATAAACGTGAGATGATTCGATATCTTTATAACCATCAGAAC 498

Qy        502 AAGGATGGTGGGTGGGGTTTGCATATTGAAGGTCCAAGCACCATGTTTGGCTCTGTCTTG 561
              |  |||||||||||||||||||||||||||||| | |   ||  ||||||||||| ||||  |||
Db        499 AGTGATGGTGGGTGGGGTTTGCATATTGAAAGCCACAGTACCATGTTTGGTTCTGCTTTG 558

Qy        562 AGTTATATTACTCTGAGATTGCTAGGTGAGGGGCCTAATGATGGACAAGGGGAAATGGAG 621
              ||  ||| ||||||||||||||||  ||  ||  ||  ||||||||||| | |||| ||||||
Db        559 AGCTATGTTACTCTGAGATTGCTTGGGGAAGGAGCTAATGATGGAGATGGGGCAATGGAA 618

Qy        622 AAGGCACGTGACTGGATTCTAGGGCATGGTGGTGCTACTTATATAACGTCATGGGGGAAG 681
              ||||  |||  | ||||||||||  ||||||  ||||  ||    ||  ||||||||||||||
Db        619 AAGGGCCGTAAATGGATTCTAGATCATGGTAGTGCCACGGCCATTACGTCATGGGGGAAA 678

Qy        682 ATGTGGCTTTCAGTACTTGGAGTGTATGAATGGTCTGGAAATAATCCCCTGCCCCCTGAG 741
              |||||||| |||||  |||||||| | |||  |||||||| ||  |||||  ||  ||  |||||
Db        679 ATGTGGCTCTCAGTGCTTGGAGTATTTGACTGGTCTGGGAACAATCCACTCCCTCCTGAA 738

Qy        742 ATATGGCTCCTTCCATACATGCTTCCATTTCATCCAGGAAGGATGTGGTGTCACTGCCGG 801
              |||||||| |||||  ||  ||  ||||||  |  |||||  |||||  ||||||||  ||  ||||||
Db        739 ATATGGCTTCTTCCCTATATCCTTCCAATCCATCCTGGAAGAATGTGGTGCCATTGCCGG 798

Qy        802 ATGGTCTATTTGCCGATGTCCTACTTATATGGCAAGAGGTTTGTTGGTCCAATCTCACCA 861
              |||||  |||  |  ||  ||||  ||||  |  |||||  ||  |  ||||||||||  |||  |  |||
Db        799 ATGGTTTATCTTCCTATGTGCTACCTTTATGGTAAACGCTTTGTTGGTCCTATCACGCCA 858

Qy        862 ACAGTATTATCTTTGAGAAAAGAGCTTTATACAGTACCATACCATGATATAGATTGGGAT 921
              |||||  |||||||||||||||  |||||  |  |||  ||  ||  ||  ||||    ||||||  ||
Db        859 ACAGTTTTATCTTTGAGAAAGGAGCTCTTTACGGTCCCGTATCATGGAGTAGATTTGGGA 918
```

```
Qy        922 CAGGCTCGCAATTTGTGTGCAAAGGAAGATTTGTACTATC 961
              | |||||||||  ||||||||||||||||  | |||||  |
Db        919 TAAGCTCGCAATGAGTGTGCAAAGGAAGACCTTTACTACC 958
```

Search completed: November 5, 2005, 12:09:18
Job time : 8086 secs

```
Qy       2019 GGTGTATTCAAATCTGGAAGGCAACAGGTCTCATGTGGTCAACACTGGGTGGGCTATGTT 2078
              ||||| ||| ||||| | |||  |||||||||||  | |||||||| |||| |||| |
Db        482 GGTGTCTTCCAATCTTGCAGGGCACAGGTCTCATGGTGGCAACACTGGATGGGGTATGGT 541

Qy       2079 GGCTCTCATTGATGCTGGACAGGCTAAGAGAGATTCGCAACCACTGCACCG 2129
              |||||||||||||||| ||| |||| ||||||||||||||||||||| ||||
Db        542 GGCTCTCATTGATGCTTGACGGGCTGAGAGAGATTCGCAACCACTGGACCG 592
```

RESULT 14
BM527876

```
LOCUS       BM527876                 558 bp    mRNA    linear   EST 06-JUL-2004
DEFINITION  sal66h03.y1 Gm-c1061 Glycine max cDNA clone SOYBEAN CLONE ID:
            Gm-c1061-4205 5' similar to TR:Q9SXV6 Q9SXV6 CYCLOARTENOL SYNTHASE.
            ;, mRNA sequence.
ACCESSION   BM527876
VERSION     BM527876.1  GI:18733901
KEYWORDS    EST.
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 558)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V.,
            Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
            Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,
            Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,
            Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
            McCann,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            When it has been determined, an EST from the other end of this
            clone is listed in the 'Other ESTs on clone' field. This clone is
            available through: Biogenetic Services, 801 32nd Ave. Brookings, SD
            57006 USA (phone: 800 423 4163; email: info@biogeneticservices.com)
            Seq primer: -40RP from Gibco
            High quality sequence stop: 422.
FEATURES             Location/Qualifiers
     source          1. .558
                     /organism="Glycine max"
                     /mol_type="mRNA"
                     /cultivar="Raiden"
                     /db_xref="taxon:3847"
                     /clone="SOYBEAN CLONE ID: Gm-c1061-4205"
                     /tissue_type="mature flowers of field grown plants"
                     /lab_host="DH10B"
                     /clone_lib="Gm-c1061"
                     /note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
                     XhoI; The cDNA library was constructed from mRNA isolated
                     from mature flowers of field grown plants for the cultivar
                     Raiden. Complementary DNA was synthesized from mRNA using
                     a primer consisting of a poly(dT) sequence with a XhoI
                     restriction site. EcoRI adapters were ligated to the
                     blunt-ended cDNA fragments followed by XhoI digestion. The
                     cDNA fragments were directionally cloned into the
                     EcoRI-XhoI restriction site of the pBluescript vector. The
                     ligated cDNA fragments were transformed into DH10B host
                     cells (GibcoBRL). This library was constructed in the
                     laboratory of Dr. Randy Shoemaker."
ORIGIN
```

```
  Query Match             21.8%;  Score 557;  DB 4;  Length 558;
  Best Local Similarity   100.0%;  Pred. No. 5e-139;
  Matches  557;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1168 CCAAATTCTGAGGCCTTCAAGTTGCATCTTCCCAGGATTTATGATTATCTATGGATTGCA 1227
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCAAATTCTGAGGCCTTCAAGTTGCATCTTCCCAGGATTTATGATTATCTATGGATTGCA 60

Qy       1228 GAAGATGGCATGAAAATGCAGGGCTACAATGGAAGTCAACTATGGGACACTGCTTTTGCT 1287
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GAAGATGGCATGAAAATGCAGGGCTACAATGGAAGTCAACTATGGGACACTGCTTTTGCT 120

Qy       1288 GTCCAAGCAATTATTGCATCTAACCTCATTGAAGAATTTGGTCCAACTATAAGAAAAGCT 1347
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GTCCAAGCAATTATTGCATCTAACCTCATTGAAGAATTTGGTCCAACTATAAGAAAAGCT 180

Qy       1348 CATACCTATATTAAGAATTCACAGGTTTTAGAAGATTGTCCAGGTGATCTTAATAAATGG 1407
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CATACCTATATTAAGAATTCACAGGTTTTAGAAGATTGTCCAGGTGATCTTAATAAATGG 240

Qy       1408 TACCGTCACATTTCAAAAGGTGCTTGGCCTTTTTCAACTGGAGATCATGGATGGCCAATT 1467
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TACCGTCACATTTCAAAAGGTGCTTGGCCTTTTTCAACTGGAGATCATGGATGGCCAATT 300

Qy       1468 TCTGACTGCACAGCTGAAGGACTGAAAGCTGTTCTATTACTATCCAAAATTGCACCAGAA 1527
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCTGACTGCACAGCTGAAGGACTGAAAGCTGTTCTATTACTATCCAAAATTGCACCAGAA 360

Qy       1528 ATAGTTGGTGAGCCAATAGACGTGAAGCGATTATATGATTCTGTAAATGTCATTCTCTCA 1587
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ATAGTTGGTGAGCCAATAGACGTGAAGCGATTATATGATTCTGTAAATGTCATTCTCTCA 420

Qy       1588 CTACAGAATGAAGATGGTGGTTTTGCAACATATGAGCTTAAACGATCTTATAATTGGTTG 1647
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTACAGAATGAAGATGGTGGTTTTGCAACATATGAGCTTAAACGATCTTATAATTGGTTG 480

Qy       1648 GAGATAATCAATCCTGCTGAAACTTTTGGTGACATCGTTATTGATTATCCTTATGTGGAA 1707
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GAGATAATCAATCCTGCTGAAACTTTTGGTGACATCGTTATTGATTATCCTTATGTGGAA 540

Qy       1708 TGTACATCAGCAGCGAT 1724
              |||||||||||||||||
Db        541 TGTACATCAGCAGCGAT 557
```

RESULT 15
CK251308

```
LOCUS       CK251308                 958 bp    mRNA    linear   EST 30-JUL-2004
DEFINITION  EST734945 potato callus cDNA library, normalized and full-length
            Solanum tuberosum cDNA clone POCBM46 5' end, mRNA sequence.
ACCESSION   CK251308
VERSION     CK251308.1  GI:39804211
KEYWORDS    EST.
SOURCE      Solanum tuberosum (potato)
  ORGANISM  Solanum tuberosum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; lamiids; Solanales; Solanaceae; Solanum.
REFERENCE   1  (bases 1 to 958)
  AUTHORS   Buell,C.R., Hart,A., Zismann,V., Karamycheva,S.A. and Baker,B.
  TITLE     Generation of ESTs from potato callus tissue
  JOURNAL   Unpublished (2003)
COMMENT     Other_ESTs: EST734946
            Contact: Robin Buell
            The Institute for Genomic Research
            9712 Medical Center Dr, Rockville, MD 20850, USA
            Email: potato-array@tigr.org
            Clones can be requested from the University of Arizona Genomics
            Institute via http://genome.arizona.edu/orders/ .
            Seq primer: ATT TAG GTG ACA CTA TAG.
FEATURES             Location/Qualifiers
     source          1. .958
                     /organism="Solanum tuberosum"
                     /mol_type="mRNA"
                     /cultivar="Kennebec"
                     /db_xref="taxon:4113"
```

```
ORIGIN

  Query Match            21.9%;  Score 561.8;  DB 5;  Length 565;
  Best Local Similarity  99.6%;  Pred. No. 2.5e-140;
  Matches  563;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1218 ATGGATTGCAGAAGATGGCATGAAAATGCAGGGCTACAATGGAAGTCAACTATGGGACAC 1277
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGATTGCAGAAGATGGCATGAAAATGCAGGGCTACAATGGAAGTCAACTATGGGACAC 60

Qy       1278 TGCTTTTGCTGTCCAAGCAATTATTGCATCTAACCTCATTGAAGAATTTGGTCCAACTAT 1337
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGCTTTTGCTGTCCAAGCAATTATTGCATCTAACCTCATTGAAGAATTTGGTCCAACTAT 120

Qy       1338 AAGAAAAGCTCATACCTATATTAAGAATTCACAGGTTTTAGAAGATTGTCCAGGTGATCT 1397
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          121 AAGAAAAGCTCATACCTATATTAAGAATTCACAGGTTTTAGAAGATTGTCCAGGTGATCT 180

Qy       1398 TAATAAATGGTACCGTCACATTTCAAAAGGTGCTTGGCCTTTTTCAACTGGAGATCATGG 1457
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          181 TAATAAATGGTACCGTCACATTTCAAAAGGTGCTTGGCCTTTTTCAACTGGAGATCATGG 240

Qy       1458 ATGGCCAATTTCTGACTGCACAGCTGAAGGACTGAAAGCTGTTCTATTACTATCCAAAAT 1517
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          241 ATGGCCAATTTCTGACTGCACAGCTGAAGGACTGAAAGCTGTTCTATTACTATCCAAAAT 300

Qy       1518 TGCACCAGAAATAGTTGGTGAGCCAATAGACGTGAAGCGATTATATGATTCTGTAAATGT 1577
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          301 TGCACCAGAAATAGTTGGTGAGCCAATAGACGTGAAGCGATTATATGATTCTGTAAATGT 360

Qy       1578 CATTCTCTCACTACAGAATGAAGATGGTGGTTTTGCAACATATGAGCTTAAACGATCTTA 1637
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          361 CATTCTCTCACTACAGAATGAAGATGGTGGTTTTGCAACATATGAGCTTAAACGATCTTA 420

Qy       1638 TAATTGGTTGGAGATAATCAATCCTGCTGAAACTTTTGGTGACATCGTTATTGATTATCC 1697
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          421 TAATTGGTTGGAGATAATCAATCCTGCTGAAACTTTTGGTGACATCGTTATTGATTATCC 480

Qy       1698 TTATGTGGAATGTACATCAGCAGCGATTCAAGCTTTGGCATCATTTAGGAAATTATATCC 1757
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          481 TTATGTGGAATGTACATCAGCAGCGATTCAAGCTTTGGCATCATTTAGGAAATTATATCC 540

Qy       1758 TGGGCATCGCCGAGAAGAAATACAA 1782
             |||||||||||||||  |||||||
Db          541 TGGGCATCGCCGAGAAAGAATACAA 565


RESULT 13
BF068916
LOCUS       BF068916               593 bp    mRNA    linear   EST 13-JUL-2004
DEFINITION  st04f11.y1 Gm-c1065 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1065-358 5' similar to TR:Q9SXV6 Q9SXV6 CYCLOARTENOL SYNTHASE.
            ;, mRNA sequence.
ACCESSION   BF068916
VERSION     BF068916.1  GI:10845888
KEYWORDS    EST.
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 593)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V.,
            Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
            Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,
            Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,
            Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
            McCann,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            When it has been determined, an EST from the other end of this
            clone is listed in the 'Other ESTs on clone' field. This clone is
            available through: Biogenetic Services, 801 32nd Ave. Brookings, SD
            57006 USA (phone: 800 423 4163; email: info@biogeneticservices.com)
            Insert Length: 951   Std Error: 0.00
            High quality sequence stop: 388.
FEATURES             Location/Qualifiers
     source          1. .593
                     /organism="Glycine max"
                     /mol_type="mRNA"
                     /cultivar="Williams"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1065-358"
                     /tissue_type="germinating shoots"
                     /lab_host="DH10B"
                     /clone_lib="Gm-c1065"
                     /note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
                     XhoI; The cDNA library was constructed from mRNA isolated
                     germinating shoots of the cultivar Williams. The seeds
                     were allowed to germinate for 24  hours prior to being
                     cold stressed for 2 days at 4C. Complementary DNA  was
                     synthesized from mRNA using a primer consisting of a
                     poly(dT)  sequence with a XhoI restriction site. EcoRI
                     adapters were ligated to  the blunt-ended cDNA fragments
                     followed by XhoI digestion. The cDNA  fragments were
                     directionally cloned into the EcoRI-XhoI restriction  site
                     of the pBluescript vector. The ligated cDNA fragments were
                     transformed into DH10B host cells (GibcoBRL). This library
                     was  constructed in the laboratory of Dr. Randy
                     Shoemaker."
ORIGIN

  Query Match            21.9%;  Score 561.2;  DB 2;  Length 593;
  Best Local Similarity  96.8%;  Pred. No. 3.7e-140;
  Matches  572;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy       1539 GCCAATAGACGTGAAGCGATTATATGATTCTGTAAATGTCATTCTCTCACTACAGAATGA 1598
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 GCCAATAGACGTGAAGCGATTATATGATTCTGTAAATGTCATTCTCTCACTACAGAATGA 61

Qy       1599 AGATGGTGGTTTTGCAACATATGAGCTTAAACGATCTTATAATTGGTTGGAGATAATCAA 1658
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 AGATGGTGGTTTTGCAACATATGAGCTTAAACGATCTTATAATTGGTTGGAGATAATCAA 121

Qy       1659 TCCTGCTGAAACTTTTGGTGACATCGTTATTGATTATCCTTATGTGGAATGTACATCAGC 1718
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         122 TCCTGCTGAAACTTTTGGTGACATCGTTATTGATTATCCTTATGTGGAATGTACATCAGC 181

Qy       1719 AGCGATTCAAGCTTTGGCATCATTTAGGAAATTATATCCTGGGCATCGCCGAGAAGAAAT 1778
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         182 AGCGATTCAAGCTTTGGCATCATTTAGGAAATTATATCCTGGGCATCGCCGAGAAGAAAT 241

Qy       1779 ACAACATTGTATCGATAAAGCCACTACCTTCATTGAAAAAATACAAGCTTCAGATGGATC 1838
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         242 ACAACATTGTATCGATAAAGCCACTACCTTCATTGAAAAAATACAAGCTTCAGATGGATC 301

Qy       1839 ATGGTATGGTTCTTGGGGAGTTTGCTTCACTTACGGTGCTTGGTTTGGGGTAAAAGGTCT 1898
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         302 ATGGTATGGTTCTTGGGGAGTTTGCTTCACTTACGGTGCTTGGTTTGGGGTAAAAGGTNT 361

Qy       1899 GATTGCTGCTGGAAGGAGTTTCAGTAATTGCTCAAGCATCCGTAAAGCTTGTGAATTTCT 1958
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         362 GATTGCTGCTGGAAGGAGTTTCAGTAATTGCTCAAGCATGCGTAAAGCTTGTGAATTTCT 421

Qy       1959 GCTGTCCAAGCAGCTTCCTTCTGGTGGCTGGGGAGAGAGTTATCTGTCCTGTCAAAACAA 2018
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         422 GCTGTCCAAGCAGCTTCCTTCTGGTGGCTGGGGAGAGAGTTATCTGTCCTGTCGAAACAA 481
```

```
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            When it has been determined, an EST from the other end of this
            clone is listed in the 'Other ESTs on clone' field. This clone is
            available through: Biogenetic Services, 801 32nd Ave. Brookings, SD
            57006 USA (phone: 800 423 4163; email: info@biogeneticservices.com)
            Seq primer: -40RP from Gibco
            High quality sequence stop: 422.
FEATURES             Location/Qualifiers
     source          1. .569
                     /organism="Glycine max"
                     /mol_type="mRNA"
                     /cultivar="Clark"
                     /db_xref="taxon:3847"
                     /clone="SOYBEAN CLONE ID: Gm-c1049-7604"
                     /tissue_type="whole seedlings of greenhouse grown plants"
                     /dev_stage="3 week old"
                     /lab_host="DH10B"
                     /clone_lib="Gm-c1049"
                     /note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
                     XhoI; The Clark NIL was constructed and seed was provided
                     by Dr. J. Specht, University of Nebraska (Shoemaker and
                     Specht, 1995). The cDNA library was constructed from mRNA
                     isolated from whole seedlings of 3 week old greenhouse
                     grown plants. Complementary DNA was synthesized from mRNA
                     using a primer consisting of a poly(dT) sequence with a
                     XhoI restriction site and a 3' anchor. EcoRI adapters were
                     ligated to the blunt-ended cDNA fragments followed by XhoI
                     digestion. The cDNA fragments were directionally cloned
                     into the EcoRI-XhoI restriction site of the pBluescript
                     vector. The ligated cDNA fragments were transformed into
                     DH10B host cells (GibcoBRL). The library was constructed
                     in cooperation with Dr. Paul Keim's laboratory at Northern
                     Arizona University."
ORIGIN

  Query Match             22.2%;  Score 569;  DB 5;  Length 569;
  Best Local Similarity   100.0%;  Pred. No. 2.8e-142;
  Matches  569;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1412 GTCACATTTCAAAAGGTGCTTGGCCTTTTTCAACTGGAGATCATGGATGGCCAATTTCTG 1471
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCACATTTCAAAAGGTGCTTGGCCTTTTTCAACTGGAGATCATGGATGGCCAATTTCTG 60

Qy       1472 ACTGCACAGCTGAAGGACTGAAAGCTGTTCTATTACTATCCAAAATTGCACCAGAAATAG 1531
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACTGCACAGCTGAAGGACTGAAAGCTGTTCTATTACTATCCAAAATTGCACCAGAAATAG 120

Qy       1532 TTGGTGAGCCAATAGACGTGAAGCGATTATATGATTCTGTAAATGTCATTCTCTCACTAC 1591
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTGGTGAGCCAATAGACGTGAAGCGATTATATGATTCTGTAAATGTCATTCTCTCACTAC 180

Qy       1592 AGAATGAAGATGGTGGTTTTGCAACATATGAGCTTAAACGATCTTATAATTGGTTGGAGA 1651
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGAATGAAGATGGTGGTTTTGCAACATATGAGCTTAAACGATCTTATAATTGGTTGGAGA 240

Qy       1652 TAATCAATCCTGCTGAAACTTTTGGTGACATCGTTATTGATTATCCTTATGTGGAATGTA 1711
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TAATCAATCCTGCTGAAACTTTTGGTGACATCGTTATTGATTATCCTTATGTGGAATGTA 300

Qy       1712 CATCAGCAGCGATTCAAGCTTTGGCATCATTTAGGAAATTATATCCTGGGCATCGCCGAG 1771
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CATCAGCAGCGATTCAAGCTTTGGCATCATTTAGGAAATTATATCCTGGGCATCGCCGAG 360

Qy       1772 AAGAAATACAACATTGTATCGATAAAGCCACTACCTTCATTGAAAAAATACAAGCTTCAG 1831
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AAGAAATACAACATTGTATCGATAAAGCCACTACCTTCATTGAAAAAATACAAGCTTCAG 420

Qy       1832 ATGGATCATGGTATGGTTCTTGGGGAGTTTGCTTCACTTACGGTGCTTGGTTTGGGGTAA 1891
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ATGGATCATGGTATGGTTCTTGGGGAGTTTGCTTCACTTACGGTGCTTGGTTTGGGGTAA 480

Qy       1892 AAGGTCTGATTGCTGCTGGAAGGAGTTTCAGTAATTGCTCAAGCATCCGTAAAGCTTGTG 1951
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAGGTCTGATTGCTGCTGGAAGGAGTTTCAGTAATTGCTCAAGCATCCGTAAAGCTTGTG 540

Qy       1952 AATTTCTGCTGTCCAAGCAGCTTCCTTCT 1980
             |||||||||||||||||||||||||||||
Db        541 AATTTCTGCTGTCCAAGCAGCTTCCTTCT 569


RESULT 12
BM885726
LOCUS       BM885726                565 bp    mRNA    linear   EST 06-JUL-2004
DEFINITION  sam04d11.y1 Gm-c1063 Glycine max cDNA clone SOYBEAN CLONE ID:
            Gm-c1063-4414 5' similar to TR:Q9SXV6 Q9SXV6 CYCLOARTENOL SYNTHASE.
            ;, mRNA sequence.
ACCESSION   BM885726
VERSION     BM885726.1  GI:19269470
KEYWORDS    EST.
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 565)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V.,
            Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
            Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,
            Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,
            Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
            McCann,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            When it has been determined, an EST from the other end of this
            clone is listed in the 'Other ESTs on clone' field. This clone is
            available through: Biogenetic Services, 801 32nd Ave. Brookings, SD
            57006 USA (phone: 800 423 4163; email: info@biogeneticservices.com)
            Seq primer: -40RP from Gibco
            High quality sequence stop: 426.
FEATURES             Location/Qualifiers
     source          1. .565
                     /organism="Glycine max"
                     /mol_type="mRNA"
                     /cultivar="Williams"
                     /db_xref="taxon:3847"
                     /clone="SOYBEAN CLONE ID: Gm-c1063-4414"
                     /tissue_type="Germinating shoot, 24 hour germination"
                     /lab_host="DH10B"
                     /clone_lib="Gm-c1063"
                     /note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
                     XhoI; The cDNA library was constructed from mRNA isolated
                     from germinating shoots of the cultivar Williams. The
                     seeds were allowed to germinate for 24 hours prior to
                     harvesting the germinating shoots. Complementary DNA was
                     synthesized from mRNA using a primer consisting of a
                     poly(dT) sequence with a XhoI restriction site. EcoRI
                     adapters were ligated to the blunt-ended cDNA fragments
                     followed by XhoI digestion. The cDNA fragments were
                     directionally cloned into the EcoRI-XhoI restriction site
                     of the pBluescript vector. The ligated cDNA fragments were
                     transformed into DH10B host cells (Gibco BRL). This
                     library was constructed in the laboratory of Dr. Randy
                     Shoemaker at Iowa state university."
```

```
                |||||  |||||||| || || |||||||||| |||||||| || ||| | ||||| ||
Db        722 GCTGAGACTTTTGGAGATATTGTTATTGATTACCCTTATGTAGAGTGTTCCTCAGCTGCA 781

Qy       1723 ATTCAAGCTTTGGCATCATTTAGGAAATTATATCCTGGGCATCGCCGAGAAGAAATACAA 1782
                |||||||||||  || ||||||  |||||||| ||||| |||||   |||||||  |  ||
Db        782 ATTCAAGCTTTAACAGCATTTAAGAAATTATACCCTGGTCATCGTAAAGAAGAAGTTGAA 841

Qy       1783 CATTGTATCGATAAAGCCACTACCTTCATT-GAAAAAATACAAGCTTCAGATGGAT-CAT 1840
                |  |||||  | ||||||  |  |||||||| ||  || |  ||  |||||||  | | |
Db        842 CGCTGTATTGCTAAAGCTGCCGCCTTCATTGGATCAATTTTCAGAAACAGATGGCTCCCT 901

Qy       1841 GGTATGGTTCTTGGGGAGTTTGCTTCACTTACGGTGCTTGGT 1882
                |||||||  |||||||  ||||||||||||  ||  ||  ||||||
Db        902 GGTATGGCTCTTGGGCAGTTTGCTTCACCTATGGCACTTGGT 943


RESULT 10
BG648956
LOCUS       BG648956                 756 bp    mRNA    linear   EST 24-APR-2001
DEFINITION  EST510575 HOGA Medicago truncatula cDNA clone pHOGA-24I22 5' end,
            mRNA sequence.
ACCESSION   BG648956
VERSION     BG648956.1  GI:13784068
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 756)
  AUTHORS   Hahn,M.G., Ojanen-Reuhs,T., Samac,D., Town,C.D., Van Aken,S.,
            Utterback,T., Cho,J. and Fraser,C.M.
  TITLE     ESTs from roots of Medicago truncatula treated with
            oligogalacturonides of DP 6-20
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Michael G. Hahn
            Complex Carbohydrate Research Center
            University of Georgia
            220 Riverbend Road, Athens, GA 30602-4712, USA
            Tel: 706-542-4457
            Fax: 706-542-4412
            Email: hahn@ccrc.uga.edu
            G392562e TIGR sequence name: MTMCR59TK More information is
            available at: www.medicago.org
            Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
FEATURES             Location/Qualifiers
     source          1. .756
                     /organism="Medicago truncatula"
                     /mol_type="mRNA"
                     /cultivar="A17"
                     /db_xref="taxon:3880"
                     /clone="pHOGA-24I22"
                     /tissue_type="3 day old seedling roots"
                     /dev_stage="24 hours after treatment in the dark at 26 C
                     with 0.5 mg/ml oligogalacturonides (DP 6-20) in the
                     presence of 100 ug/ml Gentamicin"
                     /lab_host="XLOLR"
                     /clone_lib="HOGA"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
                     was directionally ligated into the Unizap XR vector from
                     Stratagene and packaged using Gigapack III Gold packaging
                     extracts. Plasmids containing cDNA inserts were excised
                     from the recombinant lambda-Zap phage using Ex-assist
                     helper phage and propagated in SOLR cells."
ORIGIN

  Query Match             22.3%;  Score 570.8;  DB 4;  Length 756;
  Best Local Similarity   87.7%;  Pred. No. 1e-142;
  Matches  623;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy       1605 TGGTTTTGCAACATATGAGCTTAAACGATCTTATAATTGGTTGGAGATAATCAATCCTGC 1664
                |||  |||||||||||||||| | ||||||||||  ||||||||||  |||||||||
Db          1 TGGCCTTGCAACATATGAGCTCACACGATCTTATAGCTGGTTGGAGATATGCAATCCTGC 60

Qy       1665 TGAAACTTTTGGTGACATCGTTATTGATTATCCTTATGTGGAATGTACATCAGCAGCGAT 1724
                ||||||||||| ||||| |||||||||||| |||||||||||||||||| |||||||| ||
Db         61 TGAAACTTTTGGCGACATTGTTATTGATTATACTTATGTGGAATGTACCTCAGCAGCAAT 120

Qy       1725 TCAAGCTTTGGCATCATTTAGGAAATTATATCCTGGGCATCGCCGAGAAGAAATACAACA 1784
                ||| |||||| || |||||||||| |||||||||||||||| || ||||||||||||
Db         121 TCAGGCTTTGTCAACATTTAGGAAACTATATCCTGGGCATCGTCGGGAAGAAATACAACG 180

Qy       1785 TTGTATCGATAAAGCCACTACCTTCATTGAAAAAATACAAGCTTCAGATGGATCATGGTA 1844
                 ||||| || |||||  || |||| |||||||||||||||||||||||| ||||||||
Db         181 CTGTATAGAAAAAGCTGCTGCCTTTATTGAAAAAATACAAGCTTCAGATGGTTCATGGTA 240

Qy       1845 TGGTTCTTGGGGAGTTTGCTTCACTTACGGTGCTTGGTTTGGGGTAAAAGGTCTGATTGC 1904
                |||||||||||||||||||||||||||||| ||  |||||||||||||||||| ||||||||
Db         241 TGGTTCTTGGGGAGTTTGCTTCACTTATGGCACTTGGTTTGGGGTAAAAGGGCTGATTGC 300

Qy       1905 TGCTGGAAGGAGTTTCAGTAATTGCTCAAGCATCCGTAAAGCTTGTGAATTTCTGCTGTC 1964
                ||||||  | |||||||||||||||| |||||||||||||||||||| ||| |||||||
Db         301 TGCTGGGAAGAGTTTCAGTAATTGCTTAAGCATCCGTAAAGCTTGTGACTTTTTGCTGTC 360

Qy       1965 CAAGCAGCTTCCTTCTGGTGGCTGGGGAGAGAGTTATCTGTCCTGTCAAAACAAGGTGTA 2024
                ||||||||  || |||||||| ||||||||||||||||  || ||||||||||||||||
Db         361 GAAGCAGCTCCCATCTGGTGGTTGGGGAGAGAGTTATCTTTCATGTCAAAACAAGGTGTA 420

Qy       2025 TTCAAATCTGGAAGGCAACAGGTCTCATGTGGTCAACACTGGGTGGGCTATGTTGGCTCT 2084
                |||||||  |||||  |||||||||  |||||  || ||||||||||||||||  |||||||
Db         421 TTCAAATTTGGAAAGCAACAGGTCCCATGTAGTAAACACTGGGTGGGCTATGCTGGCTCT 480

Qy       2085 CATTGATGCTGGACAGGCTAAGAGAGATTCGCAACCACTGCACCGGGCAGCTGCATACTT 2144
                ||||||  ||||  ||||||  |||||||||  |   ||||||||||| |||||||  || |||
Db         481 CATTGAGGCTGAACAGGCCAAGAGAGATCCAACACCACTGCACCGCGCAGCTGTATGCTT 540

Qy       2145 GATAAATTCCCAATTGGAGGACGGTGACTTTCCGCAGCAGGAAATAATGGGAGTCTTCAA 2204
                |||||||||  ||| |||||  | ||||||||||||  |||||||||||||||||||||
Db         541 GATAAATTCTCAAATGGAGAATGGTGACTTTCCGCAAGAGGAAATAATGGGAGTCTTCAA 600

Qy       2205 CAAGAATTGCATGATCACATACGCCGCATACAGAAACATATTCCCCATTTGGGCGTTGGG 2264
                |||  |||||||||||||||| || || ||||||||||||||||||||||  ||||||||||
Db         601 CAAAAATTGCATGATCACTTATGCTGCATACAGAAACATATTCCCCATATGGGCGTTGGG 660

Qy       2265 AGAATACCAATCCCAAGTATTGCAATCTCGTTAATCGAGCCTTAGTTAGG 2314
                ||||||||      || ||||||||  |  ||| ||     | ||||  ||||
Db         661 AGAATACCGTCGTCATGTATTGCAGACCTGTTGATTTCACTTTAGATAGG 710


RESULT 11
BU083632
LOCUS       BU083632                 569 bp    mRNA    linear   EST 02-JUL-2004
DEFINITION  sar23g10.y1 Gm-c1049 Glycine max cDNA clone SOYBEAN CLONE ID:
            Gm-c1049-7604 5' similar to TR:Q9SXV6 Q9SXV6 CYCLOARTENOL SYNTHASE.
            ;, mRNA sequence.
ACCESSION   BU083632
VERSION     BU083632.1  GI:22524821
KEYWORDS    EST.
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 569)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V.,
            Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
            Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,
            Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,
            Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
            McCann,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
```

```
Qy       1553 AGCGATTATATGATTCTGTAAATGTCATTCTCTCACTACAGAATGAAGATGGTGGTTTTG 1612
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGATTATATGATTCTGTAAATGTCATTCTCTCACTACAGAATGAAGATGGTGGTTTTG 240

Qy       1613 CAACATATGAGCTTAAACGATCTTATAATTGGTTGGAGATAATCAATCCTGCTGAAACTT 1672
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAACATATGAGCTTAAACGATCTTATAATTGGTTGGAGATAATCAATCCTGCTGAAACTT 300

Qy       1673 TTGGTGACATCGTTATTGATTATCCTTATGTGGAATGTACATCAGCAGCGATTCAAGCTT 1732
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTGGTGACATCGTTATTGATTATCCTTATGTGGAATGTACATCAGCAGCGATTCAAGCTT 360

Qy       1733 TGGCATCATTTAGGAAATTATATCCTGGGCATCGCCGAGAAGAAATACAACATTGTATCG 1792
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGGCATCATTTAGGAAATTATATCCTGGGCATCGCCGAGAAGAAATACAACATTGTATCG 420

Qy       1793 ATAAAGCCACTACCTTCATTGAAAAAATACAAGCTTCAGATGGATCATGGTATGGTTCTT 1852
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ATAAAGCCACTACCTTCATTGAAAAAATACAAGCTTCAGATGGATCATGGTATGGTTCTT 480

Qy       1853 GGGGAGTTTGCTTCACTTACGGTGCTTGGTTTGGGGTAAAAGGTCTGATTGCTGCTGGAA 1912
             |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GGGGAGATTGCTTCACTTACGGTGCTTGGTTTGGGGTAAAAGGTCTGATTGCTGCTGGAA 540

Qy       1913 GGAGTTTCAGTAATTGCTCAAGCATCCGTAAAGCTTGTGAATTTCTGCTGTCCAAGCAGC 1972
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GGAGTTTCAGTAATTGCTCAAGCATCCGTAAAGCTTGTGAATTTCTGCTGTCCAAGCAGC 600

Qy       1973 TTCCTTCTGGTGGCTGGGGAGAGAGTTATCTGTCCTGT 2010
             |||||||||||||||||||||||||||||||||||||||
Db        601 TTCCTTCTGGTGGCTGGGGAGAGAGTTATCTGTCCTGT 638
```

RESULT 9
CK265941
LOCUS CK265941 943 bp mRNA linear EST 03-AUG-2004
DEFINITION EST712019 potato abiotic stress cDNA library Solanum tuberosum cDNA clone POAC172 5' end, mRNA sequence.
ACCESSION CK265941
VERSION CK265941.1 GI:39822919
KEYWORDS EST.
SOURCE Solanum tuberosum (potato)
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; lamiids; Solanales; Solanaceae; Solanum.
REFERENCE 1 (bases 1 to 943)
AUTHORS Buell,C.R., Hart,A., Zismann,V., Karamycheva,S.A. and Baker,B.
TITLE Generation of ESTs from abiotic stressed potato tissue
JOURNAL Unpublished (2003)
COMMENT Other_ESTs: EST712020
Contact: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr, Rockville, MD 20850, USA
Email: potato-array@tigr.org
Clones can be requested from the University of Arizona Genomics Institute via http://genome.arizona.edu/orders/ .
Seq primer: ATT TAG GTG ACA CTA TAG.
FEATURES Location/Qualifiers
source 1. .943
/organism="Solanum tuberosum"
/mol_type="mRNA"
/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="POAC172"
/tissue_type="abiotic stress treated leaf and root tissue"
/lab_host="DH10B-TonA"
/clone_lib="potato abiotic stress cDNA library"
/note="Vector: pCMVSport6.1; Site_1: EcoRI; Site_2: NotI; supplier: Solanum tuberosum var. Kennebec plants were grown from cuttings on a 16hr light/8 hr dark cycle at 25 C for 3-4 weeks. Abiotic stress conditions were applied to four separate sets of plants. Set 1 involved saturation of the soil with 150 mM NaCl and tissues were harvested at following application of the salt stress (leaves: 2hr, 6hr, 12hr, 1d, 2d, and 4d; roots:2hr, 6hr, 12hr, and 2d). Set 2 were grown under the standard conditions and then were water stressed by withdrawal of further watering applications. Drought stressed plants were harvested after cessation of watering (leaves: 3d, 5d, and 7d; roots:3d and 5d). Set 3 were grown under the standard conditions and then were cold stressed by placement at 4 C. Cold stressed leaves were harvested at 2 hr, 6 hr, 12 hr, 1 d, and 4d and roots were harvested at 2 hr, 6 hr, 12 hr, 1 d, 2d. Set 4 were grown under the standard conditions and then were heat stressed by placement at 35 C. Heat stressed leaves were harvested at 2 hr, 6 hr, 12 hr, 1 d, 2d and 4d and heat-stressed roots were harvested at 6 hr, 12 hr, 1 d, and 4d. RNA was isolated from all tissues and equal RNA from each tissue and stress was pooled to construct the cDNA library. RNA sample."

ORIGIN

Query Match 22.4%; Score 574; DB 7; Length 943;
Best Local Similarity 77.0%; Pred. No. 1.5e-143;
Matches 725; Conservative 0; Mismatches 215; Indels 2; Gaps 2;

```
Qy        943 AAGGAAGATTTGTACTATCCTCACCCACTTGTACAGGATATTCTTTGGGCATCTCTACAC 1002
              ||||||||  | |||||  || |||||  ||  |||||  |||||  | ||||||  ||||  |||
Db          2 AAGGAAGACCTTTACTACCCACACCCTCTAGTACAAGATATCATGTGGGCAACTCTTCAC 61

Qy       1003 AAGTTCCTTGAGCCTATTCTGATGCATTGGCCTGGAAAAAGATTGAGGGAAAAGGCTATT 1062
              |||  |  | ||  ||||||  ||||||||||||  ||  ||||||||||||||||||  ||  |
Db         62 AAGGTTGTAGAACCTATTTTGATGCATTGGCCGGGGAAAAGATTGAGGGAAAAATCTCTG 121

Qy       1063 ATTTCTGCATTGGAGCATATACATTACGAAGATGAGAATACTCGATATATTTGCATAGGT 1122
                 |||   |||||||  ||  |||||  ||||||||  ||  |||||  ||   |  ||||||||
Db         122 CGCACTGTGATGGAGCACATTCATTATGAAGATGAAAACACTCGTTACGTATGCATAGGA 181

Qy       1123 CCTGTAAATAAGGTGTTAAATATGCTTTGCTGTTGGGTGGAAGATCCAAATTCTGAGGCC 1182
              ||||||||  |||  |  ||||||||||||||||||  |||||  ||||||||  |||||  ||  ||
Db         182 CCTGTAAACAAGATTATAAATATGCTTTGCTGCTGGGTTGAAGATCCTAATTCAGAAGCT 241

Qy       1183 TTCAAGTTGCATCTTCCCAGGATTTATGATTATCTATGGATTGCAGAAGATGGCATGAAA 1242
              ||||||||||||||||  ||  |||  ||||||||  |  |||  ||||  ||||||||||||||||
Db         242 TTCAAGTTGCATCTTCCAAGAATTCATGATTATATGTGGGTTGCTGAAGATGGCATGAAA 301

Qy       1243 ATGCAGGGCTACAATGGAAGTCAACTATGGGACACTGCTTTTGCTGTCCAAGCAATTATT 1302
              |||  ||||  ||||||||||||||||    |||||  ||  |  |||||  ||  ||||||||  |||
Db         302 ATGAAGGGGTACAATGGAAGTCAATCTTGGGATACATCATTTGCAGTTCAAGCAATCATT 361

Qy       1303 GCATCTAACCTCATTGAAGAATTTGGTCCAACTATAAGAAAAGCTCATACCTATATTAAG 1362
              ||  |      |     |||||||  |||   ||||   |  ||||||||  |||  |  |    |
Db         362 GCGACAGGGTTTGGCGAAGAATATGGATCAACATTGAGAAAAGCACATTCATTCTTATCA 421

Qy       1363 AATTCACAGGTTTTAGAAGATTGTCCAGGTGATCTTAATAAATGGTACCGTCACATTTCA 1422
              |||  ||||  ||  |||||   ||||  ||  ||||||||   ||    |||||  |||||  ||||||
Db         422 AATACACAAGTGTTAGATAATTGCCCGGGTGATCTAGATTTCTGGTATCGTCATATTTCA 481

Qy       1423 AAAGGTGCTTGGCCTTTTTCAACTGGAGATCATGGATGGCCAATTTCTGACTGCACAGCT 1482
              ||||||||||||||  ||||||||||  |||||||||  |||||  |||||  ||  ||  ||  ||
Db         482 AAAGGTGCTTGGCCCTTTTCAACTGCAGATCATGGGTGGCCCATTTCAGATTGTACTGCA 541

Qy       1483 GAAGGACTGAAAGCTGTTCTATTACTATCCAAAATTGCACCAGAAATAGTTGGTGAGCCA 1542
              ||  |||||  ||  |||||||||   ||  ||||  ||   |  ||  ||||  |||||  |||  |  |||
Db         542 GAGGGACTTAAGGCTGTTCTTCTATTATCAAAGCTACCATCAGAGATAGTCGGTAATCCA 601

Qy       1543 ATAGACGTGAAGCGATTATATGATTCTGTAAATGTCATTCTCTCACTACAGAATGAAGAT 1602
              |  ||  |  ||||||  ||  ||||||  ||||||||||   |  ||  ||   ||||||||    |  |
Db         602 TTGGAGGCGAAGCGCTTGTATGATGCTGTAAATGTTTTACTTTCTTTACAGAATAGTGGT 661

Qy       1603 GGTGGTTTTGCAACATATGAGCTTAAACGATCTTATAATTGGTTGGAGATAATCAATCCT 1662
              |||||  ||||||||||||||  ||    |  |  ||  |||    |||||||||  |||||||  |||
Db         662 GGTGGCTTTGCAACATATGAACTCTCAAGGTCATATCCATGGTTGGAGCTAATCAACCCT 721

Qy       1663 GCTGAAACTTTTGGTGACATCGTTATTGATTATCCTTATGTGGAATGTACATCAGCAGCG 1722
```

Site_2: HindIII; This library was constructed from mRNA isolated from senescing leaf tissue of mature greenhouse grown plants of the cultivar Williams. Complementary DNA was synthesized from mRNA using a 3' anchored poly(dT) primer. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by digestion with EcoRI and HindIII. The cDNA fragments were directionally cloned into the EcoRI-HindIII restriction site of the pT7T3-Pac vector. The ligated cDNA fragments were transformed into DH10B host cells. This library was constructed by Dr. Randy Shoemaker."

ORIGIN

Query Match 24.9%; Score 637.2; DB 2; Length 666;
Best Local Similarity 97.3%; Pred. No. 1e-160;
Matches 648; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

```
Qy       1369 CAGGTTTTAGAAGATTGTCCAGGTGATCTTAATAAATGGTACCGTCACATTTCAAAAGGT 1428
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAGGTTTTAGAAGATTGTCCAGGTGATCTTAATAAATGGTACCGTCACATTTCAAAAGGT 60

Qy       1429 GCTTGGCCTTTTTCAACTGGAGATCATGGATGGCCAATTTCTGACTGCACAGCTGAAGGA 1488
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCTTGGCCTTTTTCAACTGGAGATCATGGATGGCCAATTTCTGACTGCACAGCTGAAGGA 120

Qy       1489 CTGAAAGCTGTTCTATTACTATCCAAAATTGCACCAGAAATAGTTGGTGAGCCAATAGAC 1548
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTGAAAGCTGTTCTATTACTATCCAAAATTGCACCAGAAATAGTTGGTGAGCCAATAGAC 180

Qy       1549 GTGAAGCGATTATATGATTCTGTAAATGTCATTCTCTCACTACAGAATGAAGATGGTGGT 1608
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GTGAAGCGATTATATGATTCTGTAAATGTCATTCTCTCACTACAGAATGAAGATGGTGGT 240

Qy       1609 TTTGCAACATATGAGCTTAAACGATCTTATAATTGGTTGGAGATAATCAATCCTGCTGAA 1668
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTTGCAACATATGAGCTTAAACGATCTTATAATTGGTTGGACATAATCAATCCTGCTGAA 300

Qy       1669 ACTTTTGGTGACATCGTTATTGATTATCCTTATGTGGAATGTACATCAGCAGCGATTCAA 1728
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ACTTTTGGTGACATCGTGATTGATTATCCTTATGTGGAATGTACATCAGCAGCGATTCAA 360

Qy       1729 GCTTTGGCATCATTTAGGAAATTATATCCTGGGCATCGCCGAGAAGAAATACAACATTGT 1788
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCTTTGGCATCATTTAGGAAATTATATCCTGGGCATCGCCGAGAAGAAATACAACATTGT 420

Qy       1789 ATCGATAAAGCCACTACCTTCATTGAAAAAATACAAGCTTCAGATGGATCATGGTATGGT 1848
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ATCGATAAAGCCACTACCTTCATTGAAAAAATACAAGCTTCAGATGGATCATGGTATGGT 480

Qy       1849 TCTTGGGGAGTTTGCTTCACTTACGGTGCTTGGTTTGGGGTAAAAGGTCTGATTGCTGCT 1908
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TCTTGGGGAGTTTGCTCCACTTACGGTGCTTGTCTTGGGGTAAAAGGTCTGATTGCTGCT 540

Qy       1909 GGAAGGAGTTTCAGTAATTGCTCAAGCATCCGTAAAGCTTGTGAATTTCTGCTGTCCAAG 1968
             |||| ||||||||||||||||||||||| || ||||||||||| ||||||||||||||||
Db        541 GGAAAGAGTTTCAGTAATTGCTCAAGCATGCGCTAAGCTTGTGAATATCTGCTGTCCAAG 600

Qy       1969 CAGCTTCCTTCTGGTGGCTGGGGAGAGAGTTATCTGTCCTGTCAAAACAAGGTGTATTCA 2028
             |||||| |||||||||||||||| |||   |||||||||||||  |||||| |||||||||
Db        601 CAGCTTTCTTCTGGTGGCTGGGTAGATGATTATCTGTCCTGTTCAAACAATGTGTATTCA 660

Qy       2029 AATCTG 2034
             ||||||
Db        661 AATCTG 666
```

RESULT 8
BU926899
LOCUS BU926899 638 bp mRNA linear EST 01-JUL-2004
DEFINITION sas93a11.y1 Gm-c1036 Glycine max cDNA clone SOYBEAN CLONE ID: Gm-c1036-11565 5' similar to TR:Q9SXV6 Q9SXV6 CYCLOARTENOL SYNTHASE. ;, mRNA sequence.
ACCESSION BU926899
VERSION BU926899.1 GI:24136389
KEYWORDS EST.
SOURCE Glycine max (soybean)
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
REFERENCE 1 (bases 1 to 638)
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE Public Soybean EST Project
JOURNAL Unpublished (1999)
COMMENT Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
When it has been determined, an EST from the other end of this clone is listed in the 'Other ESTs on clone' field. This clone is available through: Biogenetic Services, 801 32nd Ave. Brookings, SD 57006 USA (phone: 800 423 4163; email: info@biogeneticservices.com)
Seq primer: -40RP from Gibco
High quality sequence stop: 445.

FEATURES Location/Qualifiers
source 1. .638
/organism="Glycine max"
/mol_type="mRNA"
/cultivar="Jack"
/db_xref="taxon:3847"
/clone="SOYBEAN CLONE ID: Gm-c1036-11565"
/tissue_type="somatic embryos cultured on MSD 20"
/lab_host="DH10B"
/clone_lib="Gm-c1036"
/note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; This cDNA library was constructed from mRNA isolated from somatic embryos (age ranging from 2 months to 9 months) cultured on MSD 20. The library was prepared using the Life Technologies pSuperScript cDNA library construction kit. Complementary DNA was synthesized from mRNA using a poly (dT) sequence with a NotI restrictions site. SalI linkers adapters were ligated to the blunt-ended cDNA fragments followed by NotI digestion. The cDNA fragments were directionally cloned into the NotI-SalI restriction site of the pSPORT1 vector. The ligated cDNA fragments were transformed into E.coli ElectroMax DH10B host cells. This library was constructed in the laboratory of Dr. Lila Vodkin by Anu Khanna at the University of Illinois at Urbana-Champaign. e-mail: l-vodkin@uiuc.edu"

ORIGIN

Query Match 24.9%; Score 636.4; DB 5; Length 638;
Best Local Similarity 99.8%; Pred. No. 1.7e-160;
Matches 637; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       1373 TTTTAGAAGATTGTCCAGGTGATCTTAATAAATGGTACCGTCACATTTCAAAAGGTGCTT 1432
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TTTTAGAAGATTGTCCAGGTGATCTTAATAAATGGTACCGTCACATTTCAAAAGGTGCTT 60

Qy       1433 GGCCTTTTTCAACTGGAGATCATGGATGGCCAATTTCTGACTGCACAGCTGAAGGACTGA 1492
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCCTTTTTCAACTGGAGATCATGGATGGCCAATTTCTGACTGCACAGCTGAAGGACTGA 120

Qy       1493 AAGCTGTTCTATTACTATCCAAAATTGCACCAGAAATAGTTGGTGAGCCAATAGACGTGA 1552
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AAGCTGTTCTATTACTATCCAAAATTGCACCAGAAATAGTTGGTGAGCCAATAGACGTGA 180
```

```
  TITLE     Molecular determination of soybean composition
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Basil J. Nikolau and Cyril C. Periappuram
            Department of Biochemistry, Biophysics and Molecular Biology
            Iowa State University
            2210 Molecular Biology Building, Ames, IA 50011, USA
            Tel: 515 294 9423
            Fax: 515 294 0453
            Email: dimmas@iastate.edu
            Individual base call and confidence value were assigned using the
            Phred software (http://www.phrap.org/). Overall sequence quality
            assessment and vector trimming were conducted using the Lucy
            software (http://www.tigr.org/software/).
            This clone was originally generated by the Public Soybean EST
            Project (http://129.186.26.94/soybeanest.html)/Shoemaker, R
            (rcsshoe@iastate.edu).
            This clone is available through: Biogenetic Services, 801 32nd Ave,
            Brookings, SD 57006. For further information call 605-697-8500 or
            contact info@biogeneticservices.com
            Seq primer: T7-1 (5' AAT ACG ACT CAC TAT AG 3').
FEATURES             Location/Qualifiers
     source          1. .689
                     /organism="Glycine max"
                     /mol_type="mRNA"
                     /db_xref="taxon:3847"
                     /clone="Gm-r1030-2160"
                     /lab_host="DH10B"
                     /clone_lib="Gm-r1030"
                     /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This
                     cDNA library was constructed from mRNA isolated from
                     immature cotyledons of greenhouse grown plants
                     (individual seed fresh weight of 100-300mg). The library
                     was prepared using the Life Technologies pSuperScript cDNA
                     library construction kit. Complementary DNA was
                     synthesized from mRNA using a poly(dT) sequence with a
                     NotI restriction site. SalI linkers adapters were ligated
                     to the blunt-ended cDNA fragments followed by NotI
                     digestion. The cDNA fragments were directionally cloned
                     into the NotI-SalI restriction site of the pSPORT1
                     vector. The ligated cDNA fragments were transformed into
                     E. coli ElectroMax DH10B host cells. This library was
                     constructed by Dr. Lila Vodkin and Dr. Anu Khanna. Note
                     that Gm-r1030 is a re-rack of Gm-c1007."
ORIGIN

  Query Match             26.4%;  Score 677;  DB 7;  Length 689;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-171;
  Matches  677;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1617 ATATGAGCTTAAACGATCTTATAATTGGTTGGAGATAATCAATCCTGCTGAAACTTTTGG 1676
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 ATATGAGCTTAAACGATCTTATAATTGGTTGGAGATAATCAATCCTGCTGAAACTTTTGG 72

Qy       1677 TGACATCGTTATTGATTATCCTTATGTGGAATGTACATCAGCAGCGATTCAAGCTTTGGC 1736
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 TGACATCGTTATTGATTATCCTTATGTGGAATGTACATCAGCAGCGATTCAAGCTTTGGC 132

Qy       1737 ATCATTTAGGAAATTATATCCTGGGCATCGCCGAGAAGAAATACAACATTGTATCGATAA 1796
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 ATCATTTAGGAAATTATATCCTGGGCATCGCCGAGAAGAAATACAACATTGTATCGATAA 192

Qy       1797 AGCCACTACCTTCATTGAAAAAATACAAGCTTCAGATGGATCATGGTATGGTTCTTGGGG 1856
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        193 AGCCACTACCTTCATTGAAAAAATACAAGCTTCAGATGGATCATGGTATGGTTCTTGGGG 252

Qy       1857 AGTTTGCTTCACTTACGGTGCTTGGTTTGGGGTAAAAGGTCTGATTGCTGCTGGAAGGAG 1916
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        253 AGTTTGCTTCACTTACGGTGCTTGGTTTGGGGTAAAAGGTCTGATTGCTGCTGGAAGGAG 312

Qy       1917 TTTCAGTAATTGCTCAAGCATCCGTAAAGCTTGTGAATTTCTGCTGTCCAAGCAGCTTCC 1976
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        313 TTTCAGTAATTGCTCAAGCATCCGTAAAGCTTGTGAATTTCTGCTGTCCAAGCAGCTTCC 372

Qy       1977 TTCTGGTGGCTGGGGAGAGAGTTATCTGTCCTGTCAAAACAAGGTGTATTCAAATCTGGA 2036
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        373 TTCTGGTGGCTGGGGAGAGAGTTATCTGTCCTGTCAAAACAAGGTGTATTCAAATCTGGA 432

Qy       2037 AGGCAACAGGTCTCATGTGGTCAACACTGGGTGGGCTATGTTGGCTCTCATTGATGCTGG 2096
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        433 AGGCAACAGGTCTCATGTGGTCAACACTGGGTGGGCTATGTTGGCTCTCATTGATGCTGG 492

Qy       2097 ACAGGCTAAGAGAGATTCGCAACCACTGCACCGGGCAGCTGCATACTTGATAAATTCCCA 2156
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        493 ACAGGCTAAGAGAGATTCGCAACCACTGCACCGGGCAGCTGCATACTTGATAAATTCCCA 552

Qy       2157 ATTGGAGGACGGTGACTTTCCGCAGCAGGAAATAATGGGAGTCTTCAACAAGAATTGCAT 2216
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        553 ATTGGAGGACGGTGACTTTCCGCAGCAGGAAATAATGGGAGTCTTCAACAAGAATTGCAT 612

Qy       2217 GATCACATACGCCGCATACAGAAACATATTCCCCATTTGGGCGTTGGGAGAATACCAATC 2276
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        613 GATCACATACGCCGCATACAGAAACATATTCCCCATTTGGGCGTTGGGAGAATACCAATC 672

Qy       2277 CCAAGTATTGCAATCTC 2293
             |||||||||||||||||
Db        673 CCAAGTATTGCAATCTC 689


RESULT 7
BE330835
LOCUS       BE330835                 666 bp    mRNA    linear   EST 13-JUL-2004
DEFINITION  so86a10.y1 Gm-c1041 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1041-211 5' similar to TR:O23909 O23909 CYCLOARTENOL SYNTHASE
            ;, mRNA sequence.
ACCESSION   BE330835
VERSION     BE330835.1  GI:9204611
KEYWORDS    EST.
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 666)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V.,
            Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
            Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,
            Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,
            Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
            McCann,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            When it has been determined, an EST from the other end of this
            clone is listed in the 'Other ESTs on clone' field. This clone is
            available through: Biogenetic Services, 801 32nd Ave. Brookings, SD
            57006 USA (phone: 800 423 4163; email: info@biogeneticservices.com)
            High quality sequence stop: 438.
FEATURES             Location/Qualifiers
     source          1. .666
                     /organism="Glycine max"
                     /mol_type="mRNA"
                     /cultivar="Williams"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1041-211"
                     /tissue_type="Senescing leaves, mature plant, greenhouse
                     grown"
                     /lab_host="DH10B"
                     /clone_lib="Gm-c1041"
                     /note="Vector: pT7T3Pac (Pharmacia); Site_1: EcoRI;
```

```
Db        841 CATTGGAGCTGAACAGGC 858


RESULT 5
BE821083/c
LOCUS       BE821083               729 bp    mRNA    linear   EST 24-MAY-2001
DEFINITION  GM700014A10B9 Gm-r1070 Glycine max cDNA clone Gm-r1070-5441 3',
            mRNA sequence.
ACCESSION   BE821083
VERSION     BE821083.1  GI:10253317
KEYWORDS    EST.
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 729)
  AUTHORS   Vodkin,L., Keim,P., Shoemaker,R., Retzel,E., Khanna,A., Coryell,V.,
            Erpelding,J., Raph,C., Shoop,E., Pardinas,J., Liu,L. and Lewin,H.
  TITLE     A Functional Genomics Program for Soybean (NSF 9872565)
  JOURNAL   Unpublished (1999)
COMMENT     Other_ESTs: AW459730 corresponding to Gm-c1016-7527 (5')
            Contact: Vodkin, L.O., PI, A Functional Genomics Program for
            Soybean (NSF 9872565)
            Lewin, H. A., Director, Keck Center for Comparative and Functional
            Genomics
            University of Illinois
            Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
            Tel: (217) 244-6147
            Fax: (217) 333-4582
            Email: l-vodkin@uiuc.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134. For further information
            call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
            427-3324 or contact:clones@genomesystems.com or info@genome
            ystems.com web site:www.genomesystems.com
            Seq primer: 5'-TTTTTTTTTTTTTTTTTTT(A/C/G)-3'.
FEATURES             Location/Qualifiers
     source          1. .729
                     /organism="Glycine max"
                     /mol_type="mRNA"
                     /db_xref="taxon:3847"
                     /clone="Gm-r1070-5441"
                     /clone_lib="Gm-r1070"
                     /note="The library Gm-r1070 is a sequence-driven, reracked
                     set of 9,216  clones selected from cDNA libraries from
                     various tissues and stages  of development of soybean that
                     represent 2,639 sequences from  immature cotyledons, 1,770
                     from immature seed coats, 3,938 from  flowers, and 869
                     from young pods.  The 5' ESTs of the source clones from
                     the different libraries was  used to select singletons, or
                     a representative of each contig, which  were reracked to
                     form library Gm-r1070.  The cDNA clones of the reracked
                     Gm-r1070 library were then sequenced  at the 3' end.  The
                     contig analysis to select unique genes was performed by
                     the  laboratory of Ernest Retzel, Center for Computational
                     Genomics and  Bioinformatics, University of Minnesota,
                     http://www.cbc.umn.edu/ResearchProjects/Soybean/index.html
                     .  Reracking was performed by Genome Systems, St. Louis,
                     http://www.genomesystems.com, and 3' sequencing by the
                     Keck Center  for Comparative and Functional Genomics,
                     University of Illinois,
                     http://www.life.uiuc.edu/biotech/keck.html.  Note: The
                     corresponding 5' EST from each clone in the Gm-r1070
                     library is listed in the 'OTHER EST' field.  The detailed
                     information on the source library for each clone can  also
                     be obtained by referring to the Genome Systems clone ID of
                     the  original cDNA library that is also listed under
                     'OTHER EST'."
ORIGIN

  Query Match             27.3%;  Score 698.4;  DB 2;  Length 729;
  Best Local Similarity   95.9%;  Pred. No. 3e-177;
  Matches  699;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy       1627 AAACGATCTTATAATTGGTTGGAGATAATCAATCCTGCTGAAACTTTTGGTGACATCGTT 1686
             |      |  ||     |    |||||||||||| ||||||||||||| |||||||||
Db        729 ANNNNNNCNNATNNNNNGNNNNAGATAATCAATCNTGCTGAAACTTTTGNNGACATCGTT 670

Qy       1687 ATTGATTATCCTTATGTGGAATGTACATCAGCAGCGATTCAAGCTTTGGCATCATTTAGG 1746
             |||||  ||| ||||||||||||||||||||||||||||||||||||   ||||||||||||
Db        669 ATTGANTATNCTTATGTGGAATGTACATCAGCAGCGATTCAAGCTNNNGCATCATTTAGG 610

Qy       1747 AAATTATATCCTGGGCATCGCCGAGAAGAAATACAACATTGTATCGATAAAGCCACTACC 1806
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        609 AAATTATATCCTGGGCATCGCCGAGAAGAAATACAACATTGTATCGATAAAGCCACTACC 550

Qy       1807 TTCATTGAAAAAATACAAGCTTCAGATGGATCATGGTATGGTTCTTGGGGAGTTTGCTTC 1866
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        549 TTCATTGAAAAAATACAAGCTTCAGATGGATCATGGTATGGTTCTTGGGGAGTTTGCTTC 490

Qy       1867 ACTTACGGTGCTTGGTTTGGGGTAAAAGGTCTGATTGCTGCTGGAAGGAGTTTCAGTAAT 1926
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        489 ACTTACGGTGCTTGGTTTGGGGTAAAAGGTCTGATTGCTGCTGGAAGGAGTTTCAGTAAT 430

Qy       1927 TGCTCAAGCATCCGTAAAGCTTGTGAATTTCTGCTGTCCAAGCAGCTTCCTTCTGGTGGC 1986
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        429 TGCTCAAGCATCCGTAAAGCTTGTGAATTTCTGCTGTCCAAGCAGCTTCCTTCTGGTGGC 370

Qy       1987 TGGGGAGAGAGTTATCTGTCCTGTCAAAACAAGGTGTATTCAAATCTGGAAGGCAACAGG 2046
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        369 TGGGGAGAGAGTTATCTGTCCTGTCAAAACAAGGTGTATTCAAATCTGGAAGGCAACAGG 310

Qy       2047 TCTCATGTGGTCAACACTGGGTGGGCTATGTTGGCTCTCATTGATGCTGGACAGGCTAAG 2106
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        309 TCTCATGTGGTCAACACTGGGTGGGCTATGTTGGCTCTCATTGATGCTGGACAGGCTAAG 250

Qy       2107 AGAGATTCGCAACCACTGCACCGGGCAGCTGCATACTTGATAAATTCCCAATTGGAGGAC 2166
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        249 AGAGATTCGCAACCACTGCACCGGGCAGCTGCATACTTGATAAATTCCCAATTGGAGGAC 190

Qy       2167 GGTGACTTTCCGCAGCAGGAAATAATGGGAGTCTTCAACAAGAATTGCATGATCACATAC 2226
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        189 GGTGACTTTCCGCAGCAGGAAATAATGGGAGTCTTCAACAAGAATTGCATGATCACATAC 130

Qy       2227 GCCGCATACAGAAACATATTCCCCATTTGGGCGTTGGGAGAATACCAATCCCAAGTATTG 2286
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        129 GCCGCATACAGAAACATATTCCCCATTTGGGCGTTGGGAGAATACCAATCCCAAGTATTG 70

Qy       2287 CAATCTCGTTAATCGAGCCTTAGTTAGGGTGTCATCCTAATTATTTCGACCTGTGGACAG 2346
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         69 CAATCTCGTTAATCGAGCCTTAGTTAGGGTGTCATCCTAATTATNNNNGCCTGTGGACAG 10

Qy       2347 TAAAAGAAA 2355
             |||||||||
Db          9 TAAAAGAAA 1


RESULT 6
CK768972
LOCUS       CK768972               689 bp    mRNA    linear   EST 20-FEB-2004
DEFINITION  Gm-r1030-2160 Gm-r1030 Glycine max cDNA clone Gm-r1030-2160 5',
            mRNA sequence.
ACCESSION   CK768972
VERSION     CK768972.1  GI:42723073
KEYWORDS    EST.
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 689)
  AUTHORS   Periappuram,C.C., Li,L., Wurtele,E.S., Westgate,M.E. and
            Nikolau,B.J.
```

```
Db        464 GCAACAGGTCTCATGTGGTCAACACTGGGTGGGCTATGTTGGCTCTCATTGATGCTGGAC 405

Qy       2099 AGGCTAAGAGAGATTCGCAACCACTGCACCGGGCAGCTGCATACTTGATAAATTCCCAAT 2158

Db        404 AGGCTAAGAGAGATTCGCAACCACTGCACCGGGCAGCTGCATACTTGATAAATTCCCAAT 345

Qy       2159 TGGAGGACGGTGACTTTCCGCAGCAGGAAATAATGGGAGTCTTCAACAAGAATTGCATGA 2218

Db        344 TGGAGGACGGTGACTTTCCGCAGCAGGAAATAATGGGAGTCTTCAACAAGAATTGCATGA 285

Qy       2219 TCACATACGCCGCATACAGAAACATATTCCCCATTTGGGCGTTGGGAGAATACCAATCCC 2278

Db        284 TCACATACGCCGCATACAGAAACATATTCCCCATTTGGGCGTTGGGAGAATACCAATCCC 225

Qy       2279 AAGTATTGCAATCTCGTTAATCGAGCCTTAGTTAGGGTGTCATCCTAATTATTTCGACCT 2338

Db        224 AAGTATTGCAATCTCGTTAATCGAGCCTTAGTTAGGGTGTCATCCTAATTATTTCGACCT 165

Qy       2339 GTGGACAGTAAAAGAAATAATAATAATAACCTATTTTTCTTTTATTTCCATGGCTCTCTT 2398

Db        164 GTGGACAGTAAAAGAAATAATAATAATAACCTATTTTTCTTTTATTTCCATGGCTCTCTT 105

Qy       2399 AAAATGTTTGTGACTAATGAGTTTAGTAGTCAGCT-AAAAAAAAAAAGCAAACACGTGGAG 2457

Db        104 AAAATGTTTGTGACTAATGAGTTTAGTAGTCAGCTAAAAAAAAAAAGCAAACACGTGGAG 45

Qy       2458 AATGCCTGTAAGTTTTTTCTATTACTCATAGACGCCTCTCTTGC 2501

Db         44 AATGCCTGTAAGTTTTTTCTATTACTCATAGACGCCTCTCTTGC 1
```

RESULT 4
BG588039

```
LOCUS       BG588039                859 bp    mRNA    linear   EST 11-APR-2001
DEFINITION  EST489814 MHAM Medicago truncatula/Glomus versiforme mixed EST
            library cDNA clone pMHAM-51P20 5' end, mRNA sequence.
ACCESSION   BG588039
VERSION     BG588039.1  GI:13603103
KEYWORDS    EST.
SOURCE      Medicago truncatula/Glomus versiforme mixed EST library
  ORGANISM  Medicago truncatula/Glomus versiforme mixed EST library
            Eukaryota; mixed EST libraries.
REFERENCE   1  (bases 1 to 859)
  AUTHORS   Harrison,M.J., Liu,J., Town,C.D., Van Aken,S., Utterback,T., Cho,J.
            and Fraser,C.M.
  TITLE     ESTs from roots of Medicago truncatula after colonization with
            Glomus versiforme, 2001
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Harrison M.J.
            Plant Biology Division
            The Samuel Roberts Noble Foundation
            2510 Sam Noble Parkway, Ardmore, OK 73401
            Tel: 580-223-5810
            Fax: 580-221-7380
            Email: mjharrison@noble.org
            Noble EST name: N387112e TIGR sequence name: MTDDN94TK More
            information is available at: http://www.medicago.org
            Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
FEATURES             Location/Qualifiers
     source          1. .859
                     /organism="Medicago truncatula/Glomus versiforme mixed EST
                     library"
                     /mol_type="mRNA"
                     /cultivar="Medicago truncatula genotype A17"
                     /db_xref="taxon:119092"
                     /clone="pMHAM-51P20"
                     /tissue_type="roots colonized with Glomus versiforme"
                     /dev_stage="Roots harvested at 10, 17, 22, 31 and 38 days
                     post-inoculation with Glomus versiforme. The library was
                     made from a mixture of RNA from each of these stages."
                     /lab_host="E. coli strain XLOLR"
                     /clone_lib="MHAM"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; cDNA was prepared from polyA+ enriched RNA from
                     roots harvested at 10, 17, 22, 31 and 38 days
                     post-inoculation with Glomus versiforme. The cDNA was
                     directionally ligated into the Unizap XR vector from
                     Stratagene and packaged using Gigapack III Gold packaging
                     extracts. Plasmids containing cDNA inserts were excised
                     from the recombinant lambda-Zap phage using Ex-assist
                     helper phage and propagated in XLOLR cells."
ORIGIN
```

Query Match 27.4%; Score 702; DB 4; Length 859;
Best Local Similarity 89.4%; Pred. No. 3.3e-178;
Matches 767; Conservative 0; Mismatches 90; Indels 1; Gaps 1;

```
Qy       1246 CAGGGCTACAATGGAAGTCAACTATGGGACACTGCTTTTGCTGTCCAAGCAATTATTGCA 1305

Db          1 CAGGGCTACAATGGCAGTCAACTATGGGATACTTCATTTATTGCCCAAGCAATTATTTCA 60

Qy       1306 TCTAACCTCATTGAAGAATTTGGTCCAACTATAAGAAAAGCTCATACCTATATTAAGAAT 1365

Db         61 ACTAACCTAATTGAAGAATTTGGTCCAACTCTAAGAAAGGCTCATACATTCATTAAGAAT 120

Qy       1366 TCACAGGTTTTAGAAGATTGTCCAGGTGATCTTAATAAATGGTACCGTCACATTTCAAAA 1425

Db        121 TCACAGGTTTCAGAAGATTGTCCTGGTGATCTTAGTAAATGGTACCGTCACATTTCAAAA 180

Qy       1426 GGTGCTTGGCCTTTTTCAACTGGAGATCATGGATGGCCAATTTCTGACTGCACAGCTGAA 1485

Db        181 GGTGCTTGGCCCTTTTCAACTGCAGATCATGGATGGCCAATATCTGACTGCACAGCTGAA 240

Qy       1486 GGACTGAAAGCTGTTCTATTACTATCCAAAATTGCACCAGAAATAGTTGGTGAGCCAATA 1545

Db        241 GGACTGAAAGCTGTCCTATTGCTATCTAAAATTGCACCTGAGATTGTTGGTGAACCATTG 300

Qy       1546 GACGTGAAGCGATTATATGATTCTGTAAATGTCATTCTCTCACTACAGAATGAAGATGGT 1605

Db        301 GATGCAAAGAGATTATATGATGCTGTAAATGTCATTATCTCATTACAGAATGAAGATGGT 360

Qy       1606 GGTTTTGCAACATATGAGCTTAAACGATCTTATAATTGGTTGGAGATAATCAATCCTGCT 1665

Db        361 GGCCTTGCAACATATGAGCTCACACGATCTTATAGCTGGTTGGAGATATGCAATCCTGCT 420

Qy       1666 GAAACTTTTGGTGACATCGTTATTGATTATCCTTATGTGGAATGTACATCAGCAGCGATT 1725

Db        421 GAAACTTTTGGCGACATTGTTATTGATTATACTTATGTGGAATGTACCTCAGCAGCAATT 480

Qy       1726 CAAGCTTTGGCATCATTTAGGAAATTATATCCTGGGCATCGCCGAGAAGAAATACAACAT 1785

Db        481 CAGGCTTTGTCAACATTTAGGAAACTATATCCTGGGCATCGTCGGGAAGAAATACAACGC 540

Qy       1786 TGTATCGATAAAGCCACTACCTTCATTGAAAAAATACAAGCTTCAGATGGATCATGGTAT 1845

Db        541 TGTATAGAAAAAGCTGCTGCCTTTATTGAAAAAATACAAGCTTCAGATGGTTCATGGTAT 600

Qy       1846 GGTTCTTGGGGAGTTTGCTTCACTTACGGTCCTTGGTTTGGGGTAAAAGGTCTGATTGCT 1905

Db        601 GGTTCTTGGGGAGTTTGCTTCACTTATGGCACTTGGTTTGGGGTAAAAGGGCTGATTGCT 660

Qy       1906 GCTGGAAGGAGTTTCAGTAATTGCTCAAGCATCCGTAAAGCTTGTGAATTTCTGCTGTCC 1965

Db        661 GCTGGGAAGAGTTTCAGTAATTGCTTAAGCATCCGTAAAGCTTGTGACTTTTTGCTGTCG 720

Qy       1966 AAGCAGCTTCCTTCTGGTGGCT-GGGGAGAGAGTTATCTGTCCTGTCAAAACAAGGTGTA 2024

Db        721 AAGCAGCTCCCATCTGGTGGTTGGGGGAGAGAGTTATCTTTCATGTCAAAACAAGGTGTA 780

Qy       2025 TTCAAATCTGGAAGGCAACAGGTCTCATGTGGTCAACACTGGGTGGGCTATGTTGGCTCT 2084

Db        781 TTCAAATTTGGAAAGCCACAGGTCCCATGTAGTAAACACTGGGTGGGCTATGCTGGCTCT 840

Qy       2085 CATTGATGCTGGACAGGC 2102
```

```
Db       1683 TTCAGCAGCAATTCAGGCCCTGACAGCATTTAAAAAGCTCTACCCTGGACACCGCAAGAG 1742

Qy       1773 AGAAATACAACATTGTATCGATAAAGCCACTACCTTCATTGAAAAAATACAAGCTTCAGA 1832
              |||||| |  | |||||    |||||  ||| ||| |||||    || |||       ||
Db       1743 TGAAATAGACAACTGTATAAGCAAAGCTGCTAGCTTTATTGAGGGTATTCAAAAAAGCGA 1802

Qy       1833 TGGATCATGGTATGGTTCTTGGGGAGTTTGCTTCACTTACGGTGCTTGGTTTGGGGTAAA 1892
              ||| |||||||||||||||||||  ||||| || || || ||  | ||||||||| |||||
Db       1803 TGGTTCATGGTATGGTTCTTGGGCTGTTTGTTTTACCTATGGCACATGGTTTGGTGTAAA 1862

Qy       1893 AGGTCTGATTGCTGCTGGAAGGAGTTTCAGTAATTGCTCAAGCATCCGTAAAGCTTGTGA 1952
              ||  |  |||||||||| |||| |||| ||  |  |    ||| | || || |||||
Db       1863 GGGATTAGTTGCTGCTGGTAGGACATTCAAAAACAGTCCTGCAATCAGAAAGGCATGTGA 1922

Qy       1953 ATTTCTGCTGTCCAAGCAGCTTCCTTCTGGTGGCTGGGGAGAGAGTTATCTGTCCTGTCA 2012
              ||| || |||| ||  |||||||||||| |||||||||| || ||| |||| |  ||
Db       1923 CTTTTTGTTGTCAAAAGAGCTTCCTTCTGGAGGCTGGGGAGAAAGCTATTTGTCATCCCA 1982

Qy       2013 AAACAAGGTGTATTCAAATCTGGAAGGCAACAGGTCTCATGTGGTCAACACTGGGTGGGC 2072
              | |  |||| ||| | ||||| ||||| ||  |  |||||| ||| |||||||| |||||
Db       1983 AGATCAGGTTTATACCAATCTCGAAGGGAAGCGACCTCATGCGGTGAACACTGGTTGGGC 2042

Qy       2073 TATGTTGGCTCTCATTGATGCTGGACAGGCTAAGAGAGATTCGCAACCACTGCACCGGGC 2132
              ||| | || || || ||||| || |||||| |||||||| |    ||  |||| || ||
Db       2043 CATGCTAGCCCTAATCGATGCAGGGCAGGCTGAGAGAGATCCAATTCCTTTGCATCGAGC 2102

Qy       2133 AGCTGCATACTTGATAAATTCCCAATTGGAGGACGGTGACTTTCCGCAGCA 2183
              |||       ||||| || |  |||| ||| || ||||| ||||| |||||
Db       2103 AGCGAAGGTTTTGATCAACTTACAATCGGAAGATGGTGAATTTCCCCAGCA 2153
```

RESULT 3
CO984997/c

```
LOCUS       CO984997                824 bp    mRNA    linear   EST 13-SEP-2004
DEFINITION  GM89023A1B05.r1 Gm-r1089 Glycine max cDNA clone Gm-r1089-8505 3,
            mRNA sequence.
ACCESSION   CO984997
VERSION     CO984997.1  GI:51344264
KEYWORDS    EST.
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 824)
  AUTHORS   Vodkin,L., Shoemaker,R., Keim,P., Polacco,J.C., Retzel,E.,
            Khanna,A., Shealy,R., Clough,S., Thibaud-Nissen,F., Coryell,V.,
            Erpelding,J., Gonzalez,D.O., Stromvik,M., Rodriguez-Huete,A.M.,
            Schweitzer,P., Gong,G. and Liu,L.
  TITLE     A Functional Genomics Program for Soybean (NSF 9872565) (2004)
  JOURNAL   Unpublished (2004)
COMMENT     Other_ESTs: BI973054 corresponding to Gm-c1084-942 (5')
            Contact: Vodkin, L.O., PI, A Functional Genomics Program for
            Soybean (NSF 9872565)
            Lewin, H. A., Director, Keck Center for Comparative and Functional
            Genomics
            University of Illinois
            Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
            Tel: (217) 244-6147
            Fax: (217) 333-4582
            Email: l-vodkin@uiuc.edu
            Plate: GM89023A1  row: B  column: 05
            Seq primer: 5'-TTTTTTTTTTTTTTTTTTT(A/C/G)-3'
            High quality sequence stop: 824.
FEATURES             Location/Qualifiers
     source          1. .824
                     /organism="Glycine max"
                     /mol_type="mRNA"
                     /db_xref="taxon:3847"
                     /clone="Gm-r1089-8505"
                     /clone_lib="Gm-r1089"
                     /note="The library Gm-r1089 is a sequence-driven, reracked
                     set of 9,216 low redundancy clones selected from 38
                     different cDNA libraries constructed from various tissues
                     and stages of development of soybean including 973 cDNAs
                     from germinating cotyledons(source library Gm-c1069,
                     Gm-c1076, and Gm-c1077); 1,465 cDNAs from various tissue
                     and organ systems of the adult plant; 476 cDNAs from adult
                     stem tissue (source library Gm-c1062); 1340 cDNAs from
                     tissue culture derived somatic embryos (source libraries
                     Gm-c1036 and Gm-c1075); 2918 cDNAs from hypocotyls or
                     young seedlings; 742 cDNAs from germinating seedlings,
                     shoot tips, or leaves exposed to various stresses(source
                     libraries Gm-c1065, Gm-c1066, and Gm-c1068); 839 cDNAs
                     from young leaves or hypocotyls exposed to bacterial and
                     fungal pathogens (source libraries Gm-c1072, Gm-c1073,
                     Gm-c1074; and Gm-c1084); and 463 from roots of young
                     plants grown in hydroponic media without phosphate (source
                     library Gm-c1087). The 5' ESTs of the source clones from
                     the different libraries were used to select singletons, or
                     a representative of each contig, which were reracked to
                     form library Gm-r1089 and the cDNA clones of the reracked
                     Gm-r1089 library were then sequenced at the 3' end. The
                     unigene selection and 3' sequencing was funded by NSF
                     Plant Genome project #9872565
                     (http://soybeangenomics.cropsci.uiuc.edu/) as part of
                     creation of a low redundancy soybean cDNA set. The source
                     cDNA libraries were constructed by the laboratories of
                     Lila Vodkin, University of Illinois, Randy Shoemaker, Iowa
                     State University, and Paul Keim, Northern Arizona
                     University as part of the Public EST project,
                     http://129.186.26.94/soybeanest.html. The contig analysis
                     to select unique genes was performed by the laboratory of
                     Ernest Retzel, Center for Computational Genomics and
                     Bioinformatics, University of Minnesota,
                     http://web.ahc.umn.edu/biodata/nsfsoy/. Reracking and 3'
                     sequencing were conducted by services of the University of
                     Illinois Keck Center for Comparative and Functional
                     Genomics http://www.biotech.uiuc.edu/keck.htm. Note: The
                     corresponding 5' EST from each clone in the Gm-r1089
                     library is listed in the 'OTHER EST' field. The detailed
                     information on the source library for each clone can also
                     be obtained by referring to the clone ID of the original
                     cDNA library that is also listed under 'OTHER EST'."
ORIGIN
```

```
  Query Match             31.7%;  Score 812;  DB 7;  Length 824;
  Best Local Similarity   99.9%;  Pred. No. 5.7e-208;
  Matches  823;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy       1679 ACATCGTTATTGATTATCCTTATGTGGAATGTACATCAGCAGCGATTCAAGCTTTGGCAT 1738
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        824 ACATCGTTATTGATTATCCTTATGTGGAATGTACATCAGCAGCGATTCAAGCTTTGGCAT 765

Qy       1739 CATTTAGGAAATTATATCCTGGGCATCGCCGAGAAGAAATACAACATTGTATCGATAAAG 1798
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         764 CATTTAGGAAATTATATCCTGGGCATCGCCGAGAAGAAATACAACATTGTATCGATAAAG 705

Qy       1799 CCACTACCTTCATTGAAAAAATACAAGCTTCAGATGGATCATGGTATGGTTCTTGGGGAG 1858
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         704 CCACTACCTTCATTGAAAAAATACAAGCTTCAGATGGATCATGGTATGGTTCTTGGGGAG 645

Qy       1859 TTTGCTTCACTTACGGTGCTTGGTTTGGGGTAAAAGGTCTGATTGCTGCTGGAAGGAGTT 1918
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         644 TTTGCTTCACTTACGGTGCTTGGTTTGGGGTAAAAGGTCTGATTGCTGCTGGAAGGAGTT 585

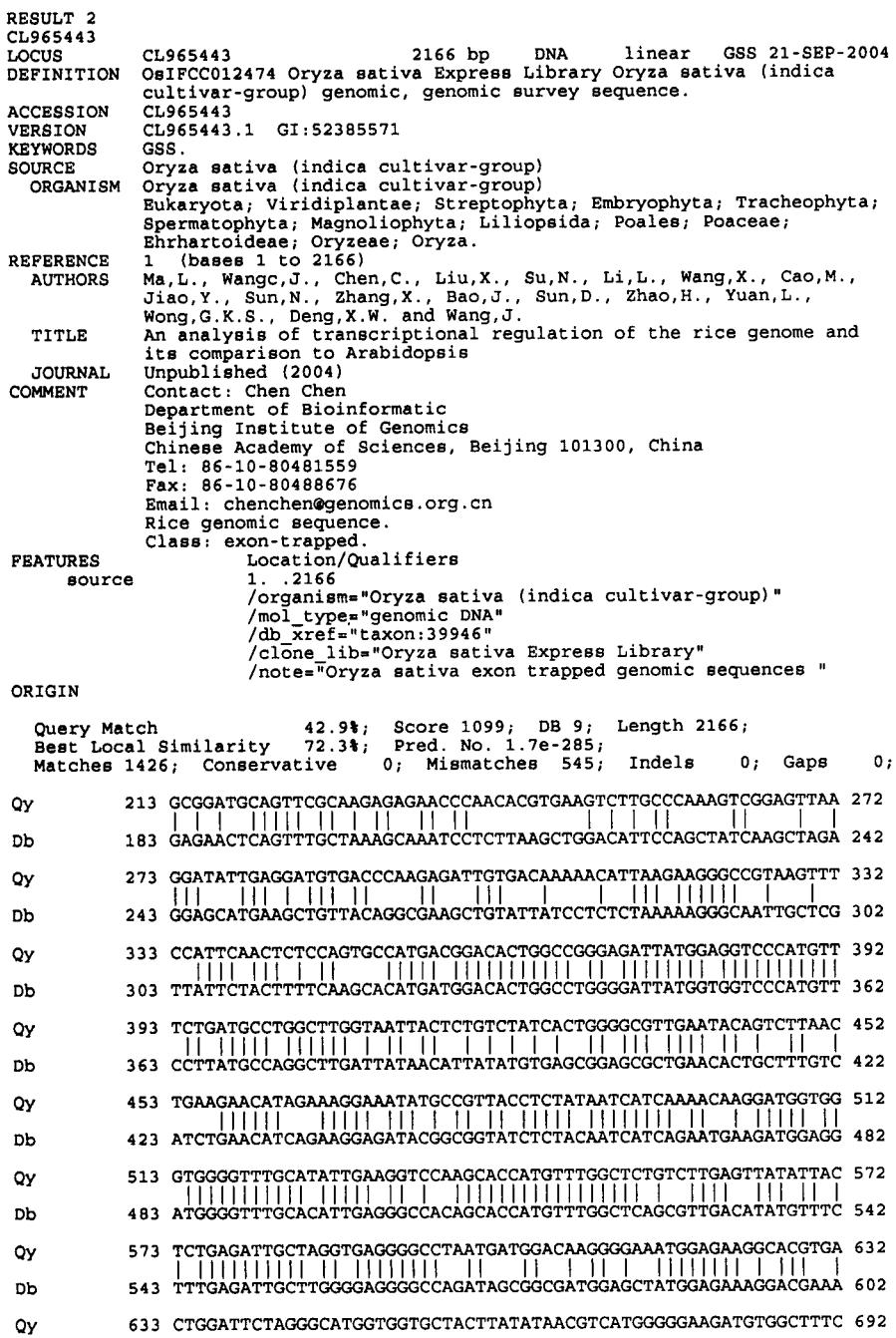
Qy       1919 TCAGTAATTGCTCAAGCATCCGTAAAGCTTGTGAATTTCTGCTGTCCAAGCAGCTTCCTT 1978
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         584 TCAGTAATTGCTCAAGCATCCGTAAAGCTTGTGAATTTCTGCTGTCCAAGCAGCTTCCTT 525

Qy       1979 CTGGTGGCTGGGGAGAGAGTTATCTGTCCTGTCAAAACAAGGTGTATTCAAATCTGGAAG 2038
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         524 CTGGTGGCTGGGGAGAGAGTTATCTGTCCTGTCAAAACAAGGTGTATTCAAATCTGGAAG 465

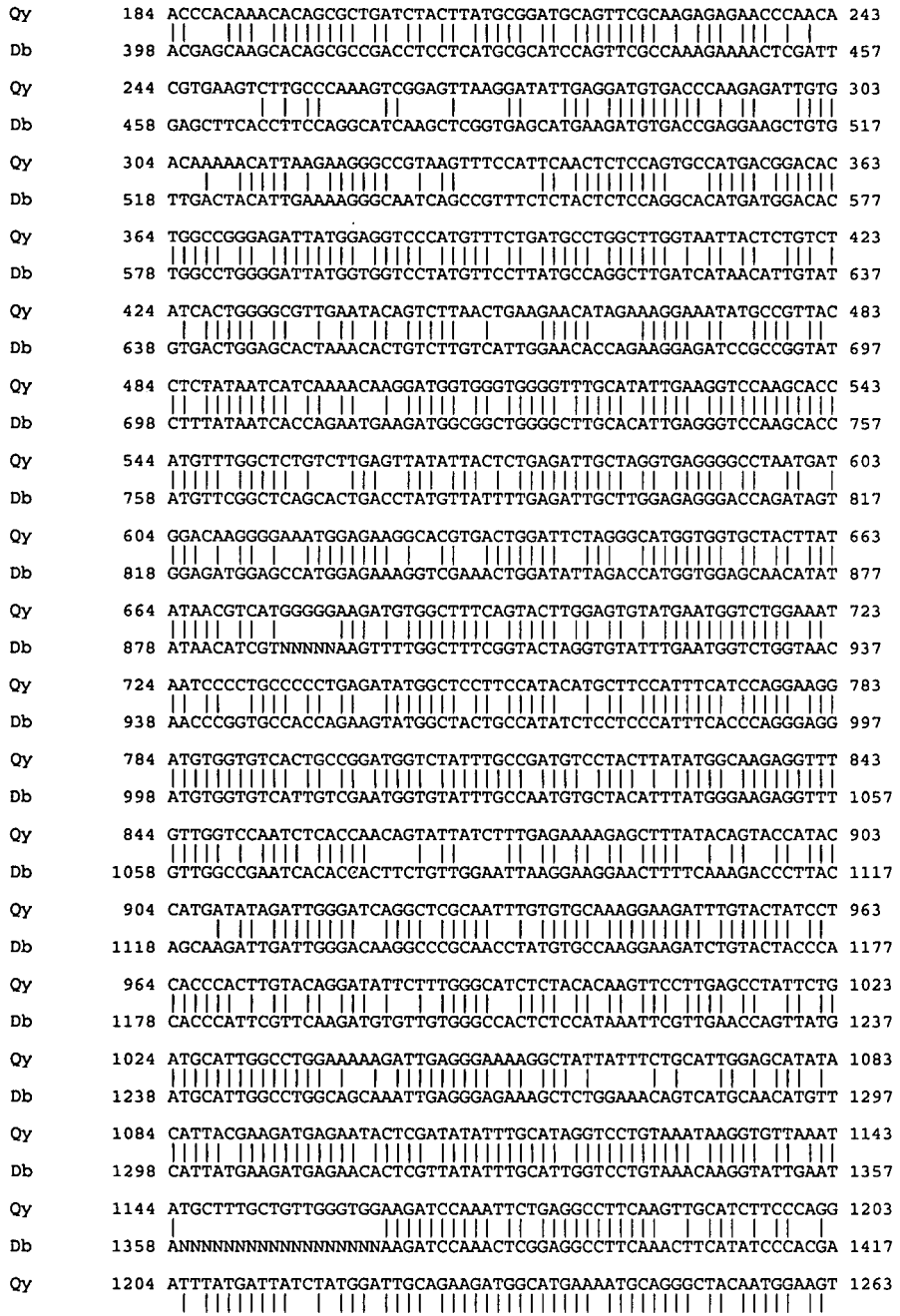
Qy       2039 GCAACAGGTCTCATGTGGTCAACACTGGGTGGGCTATGTTGGCTCTCATTGATGCTGGAC 2098
```

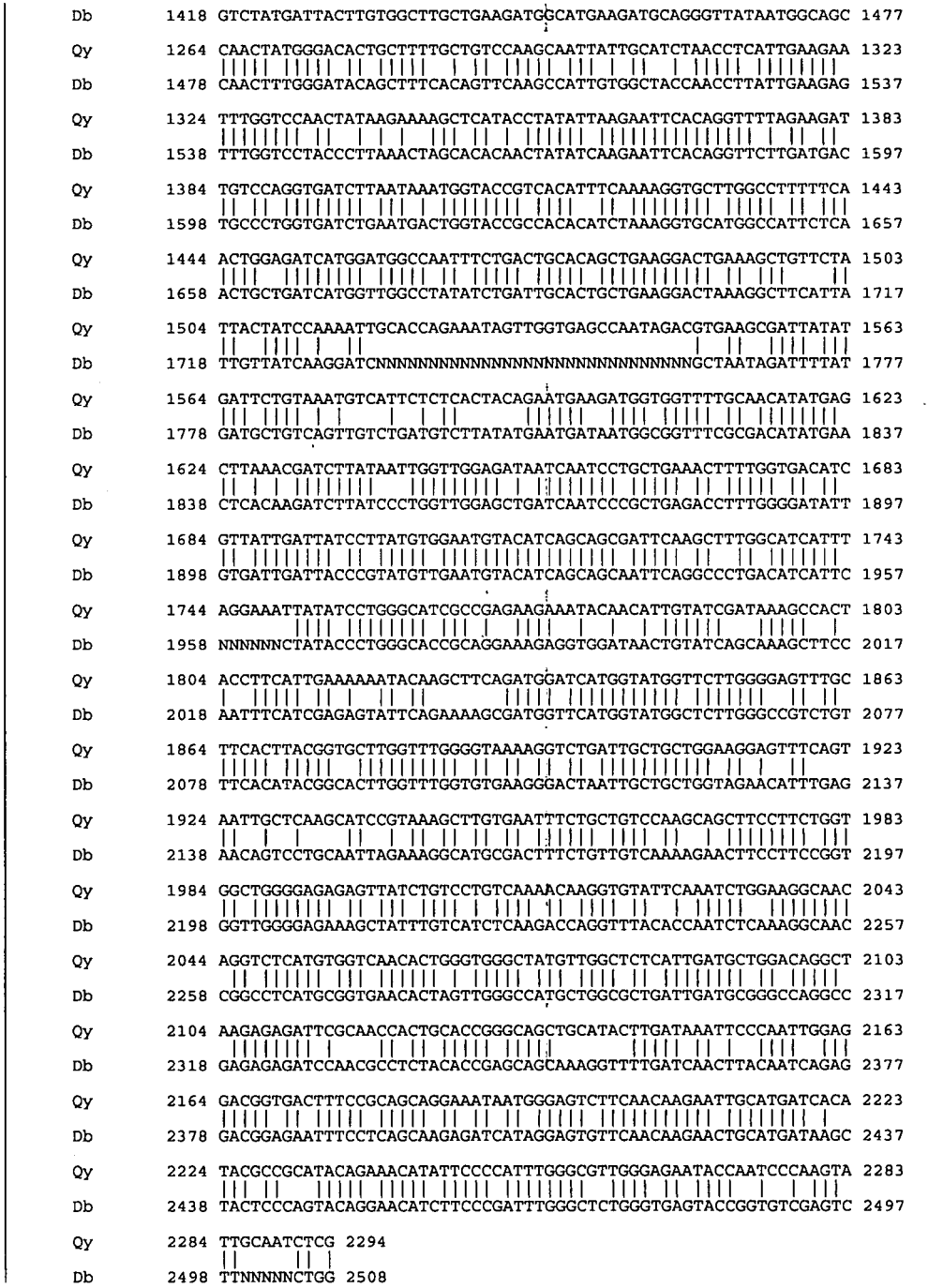
RESULT 2
CL965443
LOCUS CL965443 2166 bp DNA linear GSS 21-SEP-2004
DEFINITION OsIFCC012474 Oryza sativa Express Library Oryza sativa (indica cultivar-group) genomic, genomic survey sequence.
ACCESSION CL965443
VERSION CL965443.1 GI:52385571
KEYWORDS GSS.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 2166)
AUTHORS Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M., Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L., Wong,G.K.S., Deng,X.W. and Wang,J.
TITLE An analysis of transcriptional regulation of the rice genome and its comparison to Arabidopsis
JOURNAL Unpublished (2004)
COMMENT Contact: Chen Chen
Department of Bioinformatic
Beijing Institute of Genomics
Chinese Academy of Sciences, Beijing 101300, China
Tel: 86-10-80481559
Fax: 86-10-80488676
Email: chenchen@genomics.org.cn
Rice genomic sequence.
Class: exon-trapped.
FEATURES Location/Qualifiers
source 1. .2166
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="genomic DNA"
/db_xref="taxon:39946"
/clone_lib="Oryza sativa Express Library"
/note="Oryza sativa exon trapped genomic sequences "
ORIGIN

Query Match 42.9%; Score 1099; DB 9; Length 2166;
Best Local Similarity 72.3%; Pred. No. 1.7e-285;
Matches 1426; Conservative 0; Mismatches 545; Indels 0; Gaps 0;



Best Local Similarity 70.5%; Pred. No. 7.8e-301;
Matches 1488; Conservative 0; Mismatches 623; Indels 0; Gaps 0;





GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2005, 06:27:03 ; Search time 8073 Seconds
(without alignments)
12070.424 Million cell updates/sec

Title: US-09-419-679-5
Perfect score: 2560
Sequence: 1 ttggcctcttgccagcaaaa.....agatataaaaaaaaaaaaaa 2560

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_htc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1155.8	45.1	2920	3	AY109591	AY109591 Zea mays
	2	1099	42.9	2166	9	CL965443	CL965443 OsIFCC012
c	3	812	31.7	824	7	CO984997	CO984997 GM89023A1
	4	702	27.4	859	4	BG588039	BG588039 EST489814
c	5	698.4	27.3	729	2	BE821083	BE821083 GM700014A
	6	677	26.4	689	7	CK768972	CK768972 Gm-r1030-
	7	637.2	24.9	666	2	BE330835	BE330835 so86a10.y
	8	636.4	24.9	638	5	BU926899	BU926899 sas93a11.
	9	574	22.4	943	7	CK265941	CK265941 EST712019
	10	570.8	22.3	756	4	BG648956	BG648956 EST510575
	11	569	22.2	569	5	BU083632	BU083632 sar23g10.
	12	561.8	21.9	565	5	BM885726	BM885726 sam04d11.
	13	561.2	21.9	593	2	BF068916	BF068916 st04f11.y
	14	557	21.8	558	4	BM527876	BM527876 sal66h03.
	15	556	21.7	958	7	CK251308	CK251308 EST734945
	16	544	21.2	840	6	CB976421	CB976421 CAB40002_
	17	539	21.1	540	2	BF596965	BF596965 su82h10.y
	18	524.8	20.5	534	2	AW620615	AW620615 sj07e03.y
	19	523	20.4	594	2	BE346901	BE346901 sp32e10.y
	20	522	20.4	524	4	BG652554	BG652554 sad68f11.
	21	513.4	20.1	516	4	BM527041	BM527041 sal48h01.
	22	508	19.8	551	4	BI784597	BI784597 saf56h08.
	23	498	19.5	579	4	BI973054	BI973054 sag86d03.
	24	497	19.4	594	2	BF004071	BF004071 EST432569
	25	495.6	19.4	672	6	CD714862	CD714862 VVB205F07
	26	495.4	19.4	619	4	BG587913	BG587913 EST489688
	27	492.8	19.2	850	6	CD440985	CD440985 EL01N0563
	28	491.2	19.2	508	5	BU082808	BU082808 saq38g07.
	29	490.6	19.2	666	6	CD718769	CD718769 VVB153H11
	30	489.4	19.1	708	7	CV518867	CV518867 0089P0020
	31	487.4	19.0	803	6	CB673111	CB673111 OSJNEe07H
c	32	486.2	19.0	505	6	CD401631	CD401631 Gm_ck2400
	33	484	18.9	760	4	BI920796	BI920796 EST540719
	34	482.8	18.9	662	6	CD714635	CD714635 VVB203B10
	35	482.4	18.8	673	7	CV095432	CV095432 FAMU_USDA
	36	482.2	18.8	664	5	BU040703	BU040703 PP_LEa000
	37	479.4	18.7	707	5	BQ866177	BQ866177 QGC7E04.y
c	38	479.4	18.7	901	7	CV227637	CV227637 WS0167.B2
	39	471	18.4	768	6	CB622107	CB622107 OSIIEa08H
	40	470.2	18.4	674	4	BM437263	BM437263 VVA016H02
	41	469.4	18.3	681	6	CD006346	CD006346 VVB058B07
	42	469	18.3	831	6	CB642220	CB642220 OSJNEb02E
	43	466.6	18.2	781	6	CB651302	CB651302 OSJNEb16D
	44	456.4	17.8	770	4	BG595048	BG595048 EST493726
	45	448.4	17.5	868	7	CF511373	CF511373 CAbud0002

ALIGNMENTS

RESULT 1
AY109591
LOCUS AY109591 2920 bp mRNA linear HTC 17-OCT-2002
DEFINITION Zea mays CL201_1 mRNA sequence.
ACCESSION AY109591
VERSION AY109591.1 GI:21213371
KEYWORDS HTC.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 2920)
AUTHORS Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S., Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes
JOURNAL Unpublished (2002)
REFERENCE 2 (bases 1 to 2920)
AUTHORS Coe,E.H.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA
COMMENT If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmDB: www.zmdb.iastate.edu.
FEATURES Location/Qualifiers
source 1. .2920
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="MaizeDB:630740"
/db_xref="taxon:4577"
/clone_lib="Maize Mapping Project/DuPont Cornsensus Library"
/note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed DuPont contigs; this resource was assembled by DuPont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"
ORIGIN

Query Match 45.1%; Score 1155.8; DB 3; Length 2920;

```
Qy       1754 ATCCTGGGCATCGCCGAGAAGAAATACAACATTGTATCGATAAAGCCACTACCTTCATTG 1813
              |||| || ||| |     |||| ||| || |||  ||     || ||    |   ||| |||
Db       1762 ATCCAGGACATAGGAAGAAAGAGATAGAAAATTTCATTACCAATGCAGTTCGATTCCTTG 1821

Qy       1814 AAAAAATACAAGCTTCAGATGGATCATGGTATGGTTCTTGGGGAGTTTGCTTCACTTACG 1873
              || | | |||| |  | ||||| |||||||||||   |||||||||||||||||||| |
Db       1822 AAGATACACAAACAGCTGATGGTTCATGGTATGGAAATTGGGGAGTTTGCTTCACTTATG 1881

Qy       1874 GTGCTTGGTTTGGGGTAAAAGGTCTGATTGCTGCTGGAAGGAGTTTCAGTAATTGCTCAA 1933
              |  |||||||||    |   ||||||    |||||||| | || || ||  |||||  |
Db       1882 GCTCTTGGTTTGCACTTGGAGGTCTAGCAGCTGCTGGTAAGACTTACACCAATTGTGCTG 1941

Qy       1934 GCATCCGTAAAGCTTGTGAATTTCTGCTGTCCAAGCAGCTTCCTTCTGGTGGCTGGGGAG 1993
               ||| || |||||   | |||||||  ||  | |  ||          ||||| |||||||
Db       1942 CCATTCGCAAAGCCGTTAAATTTCTACTTACAACACAAAGAGAGGACGGTGGATGGGGAG 2001

Qy       1994 AGAGTTATCTGTCCTGTCAAAACAAGGTGTATTCAAATCTGGAAGGCAACAGGTCTCATG 2053
              |||||||||| ||  | | ||| ||| | |||  |  ||| ||||| | | | ||  |||
Db       2002 AGAGTTATCTTTCAAGCCCAAAAAAGATATATGTACCTCTAGAAGGAAGCCGATCAAATG 2061

Qy       2054 TGGTCAACACTGGGTGGGCTATGTTGGCTCTCATTGATGCTGGACAGGCTAAGAGAGATT 2113
              | ||  | || |  |||||| |  |||  || ||| ||||||||||||  | |||||
Db       2062 TTGTACATACAGCATGGGCTCTTATGGGACTAATTCATGCTGGACAGGCGGATAGAGACC 2121

Qy       2114 CGCAACCACTGCACCGGGCAGCTGCATACTTGATAAATTCCCAATTGGAGGACGGTGACT 2173
              |    || || ||||| || ||    |   | || ||||| || ||||| || ||||| |
Db       2122 CCATGCCTCTTCACCGTGCTGCAAAGTTGCTCATTAATTCTCAGTTGGAAGAGGGTGATT 2181

Qy       2174 TTCCGCAGCAGGAAATAATGGGAGTCTTCAACAAGAATTGCATGATCACATACGCCGCAT 2233
                || || |||||||| | |||||| ||||  || |||||||| |    ||  |    |
Db       2182 GGCCCCAACAGGAAATCACGGGAGTATTCATGAAAAATTGCATGTTGCATTATCCAATGT 2241

Qy       2234 ACAGAAACATATTCCCCATTTGGGCGTTGGGAGAATACCAATCCCAAGT 2282
              ||||| | || |  || || |||||  | |  ||||| | |   |  ||
Db       2242 ACAGAGATATTTATCCAATGTGGGCTCTAGCTGAATATCGAAGGCGGGT 2290
```

Search completed: November 5, 2005, 12:49:35
Job time : 1991 secs

RESULT 15
US-10-425-114-7892
; Sequence 7892, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 7892
; LENGTH: 2567
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 700725166_FLI
US-10-425-114-7892

Query Match 36.8%; Score 942.6; DB 19; Length 2567;
Best Local Similarity 64.7%; Pred. No. 1.2e-247;
Matches 1468; Conservative 0; Mismatches 789; Indels 12; Gaps 4;

```
Qy      23 GAATGTGGAAGCTCAAGTTCGCCGAAGGAGGGA---ATCCATGGCTTCGGACATTGAACA 79
           | ||||||| ||| ||| | || || ||||| |   ||||||  |   |    ||||
Db      25 GGATGTGGAGGCTGAAGATAGCAGATGGAGGAAATGATCCATACATATTCAGCACAAACA 84

Qy      80 ATCACGTTGGAAGACAGGTGTGGGAGTTCGATCCTAAGCTTGGATCGCCGCAAGATCTCC 139
           ||  |||||| || |||   |||||||| ||||||  |     ||    ||  | || |
Db      85 ATTTCGTTGGGAGGCAGACATGGGAGTTTGATCCTGAAGCAGGCAGTCCAGAGGAACGGG 144

Qy     140 TCGAGATTGAGAAAGCTCGCCAGAATTTTCACGATAACCGCTTTACCCACAAACACAGCG 199
           | || ||||   |||||| ||| ||||  || | ||||||||  |    ||| | | |||
Db     145 CCCAGGTTGAAGCAGCTCGTCAGCATTTCTACCACAACCGCTTCAAGGTCAAGCCCTGCG 204

Qy     200 CTGATCTACTTATGCGGATGCAGTTCGCAAGAGAGAACCCAACACGTGAAGTCTTGCCCA 259
           |||| || |||  |||. | ||| |    ||||| ||           ||    | ||
Db     205 CTGACCTCCTTTGGCGTTTTCAGGTTCTCAGAGAAAATAACTTCAAACAAACAATTCCTC 264

Qy     260 AAGTCGGAGTTAAGGATATTGAGGATGTGACCCAAGAGATTGTGACAAAAACATTAAGAA 319
           ||      |  | |||    |||||  | ||  |  | |  || ||||    |  | ||||
Db     265 GTGTGACTATAGAAGATGGAGAGGAAATCACATACCAAAAAGTCACAAGCGCCGTCAGAA 324

Qy     320 GGGCCGTAAGTTTCCATTCAACTCTCCAGTGCCATGACGGACACTGGCCGGGAGATTATG 379
           |||  || |    || | |  | || |||  |  ||| || || |||||  |    ||
Db     325 GGGGCGCACACCACCTTGCGGCACTGCAGACCTCTGATGGCCATTGGCCTGCTCAAATTG 384

Qy     380 GAGGTCCCATGTTTCTGATGCCTGGCTTGGTAATTACTCTGTCTATCACTGGGGCGTTGA 439
           ||||||  | ||  |  | |||   |||||  ||  | ||| ||| || ||      |
Db     385 CAGGTCCTCTCTTCTTTCTTCCTCCTTTGGTTTTTTGTATGTATATTACAGGAAATCTTG 444

Qy     440 ATACAGTCTTAACTGAAGAACATAGAAAGGAAATATGCCGTTACCTCTATAATCATCAAA 499
           |  |||| ||  | |||||||||| | || |||||     ||||||   ||| |||| || |
Db     445 AATCAGTATTTCCAGAAGAACATCGCAAAGAAATTCTTCGTTACACATATTATCACCAGA 504

Qy     500 ACAAGGATGGTGGGTGGGGTTTGCATATTGAAGGTCCAAGCACCATGTTTGGCTCTGTCT 559
           |  | || || || |||||  | || || || ||||  |||||  ||||||  |  |||
Db     505 ATGAAGACGGAGGATGGGGACTACACATAGAGGGTCATAGCACTATGTTTTGTACTGCAC 564

Qy     560 TGAGTTATATTACTCTGAGATTGCTAGGTGAGGGGCCTAATGATGGACAAG---GGGAAA 616
           |||  |||||      || || |||| || || || |||||||  || || |     |
Db     565 TGAACTATATATGCATGCGAATGCTTGGAGAAGGACCTAATGGAGGTCATGACAATGCTT 624

Qy     617 TGGAGAAGGCACGTGACTGGATTCTAGGGCATGGTGGTGCTACTTATATAACGTCATGGG 676
           |  |  ||| |  | |||||||| ||  |||||||||| ||  ||||| | |||||||
Db     625 GTGCTAGAGCAAGAAAGTGGATTCGAGATCATGGTGGTGTAACACATATACCTTCATGGG 684

Qy     677 GGAAGATGTGGCTTTCAGTACTTGGAGTGTATGAATGGTCTGGAAATAATCCCCTGCCCC 736
           | || |  |||||||  |||| || || | ||| ||||   ||||  || ||  ||||||
Db     685 GAAAAACTTGGCTTTCGATACTCGGTGTATTTGATTGGTGCGGAAGCAACCCAATGCCCC 744

Qy     737 CTGAGATATGGCTCCTTCCATACATGCTTCCATTTCATCCAGGAAGGATGTGGTGTCACT 796
           | ||| | ||| |||||||||   | |||||  | |||||||  | |||||||||| |||
Db     745 CAGAGTTTTGGATCCTTCCATCTTTTCTTCCTATGCATCCAGCTAAGATGTGGTGTTACT 804

Qy     797 GCCGGATGGTCTATTTGCCGATGTCCTACTTATATGGCAAGAGGTTTGTTGGTCCAATCT 856
           | ||  |||| ||  |||| |||||  |||||||||| ||||||||||  |||||||||
Db     805 GTCGATTGGTATACATGCCTATGTCTTACTTATATGGGAAGAGGTTTGTGGGTCCAATCA 864

Qy     857 CACCAACAGTATTATCTTTGAGAAAAGAGCTTTATACAGTACCATACCATGATATAGATT 916
           |||||   | |||   || ||| ||||| | | |||   ||| ||  |  |  | |||
Db     865 CACCACTCATCTTACAATTAAGAGAAGAGTTGTTTACTCAACCTTATGAAAAAGTTAATT 924

Qy     917 GGGATCAGGCTCGCAATTTGTGTGCAAAGGAAGATTTGTACTATCCTCACCCACTTGTAC 976
           || |  | || ||  |  ||||||||||||||| | |||||||| || ||  | |||
Db     925 GGAAGAAAGCGCGTCACCAATGTGCAAAGGAAGATCTTTACTATCCCCATCCTTTGATAC 984

Qy     977 AGGATATTCTTTGGGCATCTCTACACAAGTTCCTTGAGCCTATTCTGATGCATTGGCCTG 1036
           | ||  | | ||||    | || |||  ||| ||| ||  | || | | |||||||
Db     985 AAGACCTAATATGGGATAGTTTATACATATTCACTGAACCGCTACTTACTCGTTGGCCTT 1044

Qy    1037 GAAAAAGATTG---AGGGAAAAGGCTATTATTTCTGCATTGGAGCATATACATTACGAAG 1093
           || |  |||   || |||||||  ||      | || | ||||| ||||| ||||
Db    1045 TCAACAAGTTGATTAGAGAAAAGGCCCTTCAAGTAACTATGAAACATATTCATTATGAAG 1104

Qy    1094 ATGAGAATACTCGATATATTTGCATAGGTCCTGTAAATAAGGTGTTAAATATGCTTTGCT 1153
           ||||||  || |||||| ||   ||| |||  |||  | ||||| |||  |||||||   |
Db    1105 ATGAGACTAGTCGATACATAACCATTGGTTGTGTGGAAAAGGTTTTATGTATGCTTGCTT 1164

Qy    1154 GTTGGGTGGAAGATCCAAATTCTGAGGCCTTCAAGTTGCATCTTCCCAGGATTTATGATT 1213
           ||||||||||||||||||||    || || ||||||  |||||||  | ||| |    ||||
Db    1165 GTTGGGTGGAAGATCCAAACGGAGATGCTTTCAAGAAGCATCTTGCAAGGGTCCCAGATT 1224

Qy    1214 ATCTATGGATTGCAGAAGATGGCATGAAAATGCAGGGCTACAATGGAAGTCAACTATGGG 1273
           |  ||||| || | |||||||| ||||  ||||||  | |    ||| || |||  |||||
Db    1225 ACTTATGGGTTTCTGAAGATGGAATGACCATGCAGAGTT---TTGGTAGCCAAGAATGGG 1281

Qy    1274 ACACTGCTTTTGCTGTCCAAGCAATTATTGCATCTAACCTCATTGAAGAATTTGGTCCAA 1333
           |  |||  |||||||| |||||  |  ||||  |||||  | |||||||| |||||| |
Db    1282 ATGCTGGCTTTGCTGTTCAAGCTTTGCTTGCACTAACATAATTGAAGAAATTGGTCCTA 1341

Qy    1334 CTATAAGAAAAGCTCATACCTATATTAAGAATTCACAGGTTTTAGAAGATTGTCCAGGTG 1393
           |  |   |||||  |||  |  || ||||| || |||||     ||  ||  |  || |
Db    1342 CGTTTGCAAAAGGACATGATTTCATCAAGAAGTCTCAGGTGAAGGATAATCCTTTTGGAG 1401

Qy    1394 ATCTTAATAAATGGTACCGTCACATTTCAAAAGGTGCTTGGCCTTTTTCAACTGGAGATC 1453
           || |||| |   | | ||||| ||||| |||||  |||||  | || ||    | ||| |
Db    1402 ATTTTAAAAGTATGCATCGTCATATTTCTAAAGGGTCTTGGACATTCTCTGATCAAGACC 1461

Qy    1454 ATGGATGGCCAATTTCTGACTGCACAGCTGAAGGACTGAAAGCTGTTCTATTACTATCCA 1513
           ||||||||| | ||||||| ||||| || |||||  | ||    | |||| | ||||| |
Db    1462 ATGGATGGCAAGTTTCTGATTGCACTGCAGAAGGTTTAAAGTGTTGTCTACTTCTATCAA 1521

Qy    1514 AAATTGCACCAGAAATAGTTGGTGAGCCAATAGACGTGAAGCGATTATATGATTCTGTAA 1573
           |  ||||||| || || || ||   || ||      | ||||||| ||||| || |
Db    1522 TGTTGCCACCAGAGATTGTGGGAGAAAAGATGGAACCTGAAAGATTATACGATTCAGTCA 1581

Qy    1574 ATGTCATTCTCTCACTACAGAATGAAGATGGTGGTTTTGCAACATATGAGCTTAAACGAT 1633
           |||||  |  || || |||| | || | ||||||| ||| |||  |||| |  | ||
Db    1582 ATGTCTTGTTGTCGCTTCAGAGTAAAAAAGGTGGTTTAGCAGCATGGGAGCCTGCAGGAG 1641

Qy    1634 CTTATAATTGGTTGGAGATAATCAATCCTGCTGAAACTTTTGGTGACATCGTTATTGATT 1693
           ||  |  ||||| ||  || ||||||| | |||  |||||  |||||  || ||||
Db    1642 CTCAAGAGTGGTTAGAATTACTCAATCCCACAGAATTTTTTGCGGACATTGTAGTTGAAC 1701

Qy    1694 ATCCTTATGTGGAATGTACATCAGCAGCGATTCAAGCTTTGGCATCATTTAGGAAATTAT 1753
           ||   ||||| || || ||  | | || || |||||||| |  | || | ||| |||
Db    1702 ATGAATATGTTGAGTGCACTGGATCTGCAATCCAAGCTTTAGTTTTGTTCAAGAAGCTAT 1761
```

```
Qy       2234 ACAGAAACATATTCCCCATTTGGGCGTTGGGAGAATACCAATCCCAAGT 2282
             ||||| | || |  || || |||||  | |  ||||| | |   |  ||
Db       2255 ACAGAGATATTTATCCAATGTGGGCTCTAGCTGAATATCGAAGGCGGGT 2303


RESULT 14
US-10-607-726-3
; Sequence 3, Application US/10607726
; Publication No. US20040121439A1
; GENERAL INFORMATION:
;  APPLICANT: Falco, S. Carl
;  APPLICANT:  Famodu, Omolayo O.
;  APPLICANT:  Hitz, William D.
;  APPLICANT:  Kinney, Anthony J.
;  APPLICANT:  Rafalski, Antoni
;  APPLICANT:  McGonigle, Brian
;  APPLICANT:  Lohman, Karin
;  TITLE OF INVENTION: Enzymes Involved In Squalene Metabolizm
;  FILE REFERENCE: BB1113 US NA
;  CURRENT APPLICATION NUMBER: US/10/607,726
;  CURRENT FILING DATE:  2003-06-27
;  PRIOR APPLICATION NUMBER: US/09/419,679
;  PRIOR FILING DATE: 1999-10-15
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/105,405
;  PRIOR FILING DATE: EARLIER FILING DATE: 1998-10-23
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 3
;   LENGTH: 1882
;   TYPE: DNA
;   ORGANISM: Oryza sativa
US-10-607-726-3

  Query Match             36.9%;  Score 944;  DB 20;  Length 1882;
  Best Local Similarity   74.4%;  Pred. No. 4.1e-248;
  Matches 1190;  Conservative    0;  Mismatches  410;  Indels    0;  Gaps    0;

Qy        687 GCTTTCAGTACTTGGAGTGTATGAATGGTCTGGAAATAATCCCCTGCCCCCTGAGATATG 746
              |||||| || ||||| || | ||| |||||||| || || ||  |||| || || |||||
Db         14 GCTTTCGGTGCTTGGTGTATTTGACTGGTCTGGCAACAACCCAGTGCCACCAGAAATATG 73

Qy        747 GCTCCTTCCATACATGCTTCCATTTCATCCAGGAAGGATGTGGTGTCACTGCCGGATGGT 806
              | |  | |||||  | || ||  |||||||||| | ||||||||||| ||||||||||
Db         74 GTTGTTGCCATATTTCCTGCCGATTCATCCAGGGCGAATGTGGTGTCATTGCCGGATGGT 133

Qy        807 CTATTTGCCGATGTCCTACTTATATGGCAAGAGGTTTGTTGGTCCAATCTCACCAACAGT 866
              |||||||| |||  ||| | ||||| |||||||||| || |||  |  ||||||    |
Db *      134 TTATTTGCCTATGTGTTACATTTATGGAAAGAGGTTTGTGGGCCCAGTTACACCAATTAT 193

Qy        867 ATTATCTTTGAGAAAAGAGCTTTATACAGTACCATACCATGATATAGATTGGGATCAGGC 926
              |||    || ||||| || || ||   |||||| ||| ||||  | ||||||||  ||||
Db        194 ATTGGAATTAAGAAAGGAACTCTACGAAGTACCCTACAATGAAGTTGATTGGGACAAGGC 253

Qy        927 TCGCAATTTGTGTGCAAAGGAAGATTTGTACTATCCTCACCCACTTGTACAGGATATTCT 986
              ||||||| | ||||| |||||||||| |||||||||| || ||| | || ||||||  |  |
Db        254 TCGCAATCTATGTGCTAAGGAAGATCTGTACTATCCACATCCATTCGTGCAGGATGTATT 313

Qy        987 TTGGGCATCTCTACACAAGTTCCTTGAGCCTATTCTGATGCATTGGCCTGGAAAAAGATT 1046
              |||||  |||| |||||  ||  |||| ||    | || ||| |||||||||  || | |||
Db        314 ATGGGCCACTCTCCACAAATTTGTTGAACCAGCTATGTTGCGTTGGCCTGGGAACAAATT 373

Qy       1047 GAGGGAAAAGGCTATTATTTCTGCATTGGAGCATATACATTACGAAGATGAGAATACTCG 1106
              |||||| || ||| |     |||   || ||||||| ||||| |||||||||| || ||
Db        374 GAGGGAGAAAGCTTTGGACACTGTCATGCAGCATATTCATTATGAAGATGAGAACACCCG 433

Qy       1107 ATATATTTGCATAGGTCCTGTAAATAAGGTGTTAAATATGCTTTGCTGTTGGGTGGAAGA 1166
              |||||||||||| ||||| ||||| ||||| ||||||||||    || ||| | |||||
Db        434 ATATATTTGCATTGGTCCAGTAAACAAGGTATTAAATATGCTTGCTTGCTGGATTGAAGA 493

Qy       1167 TCCAAATTCTGAGGCCTTCAAGTTGCATCTTCCCAGGATTTATGATTATCTATGGATTGC 1226
              |||||| || ||||| |||||  | ||  |||| ||  |  | ||||| ||||||||||
Db        494 TCCAAACTCAGAGGCATTCAAACTCCACATTCCAAGAGTCCACGATTACCTATGGATTGC 553

Qy       1227 AGAAGATGGCATGAAAATGCAGGGCTACAATGGAAGTCAACTATGGGACACTGCTTTTGC 1286
              |||||||||||||||||||||||| || |||||||| || || |||||||| |||||  |
Db        554 AGAAGATGGCATGAAAATGCAGGGTTATAATGGAAGCCAGCTGTGGGACACAGCTTTCAC 613

Qy       1287 TGTCCAAGCAATTATTGCATCTAACCTCATTGAAGAATTTGGTCCAACTATAAGAAAAGC 1346
              || ||||| ||  | ||  ||  |||||||||||||||||||| ||| | | | |||
Db        614 AGTTCAAGCTATAGTGGCTACTGGCCTCATTGAAGAATTTGGTCCTACTCTTAAACTAGC 673

Qy       1347 TCATACCTATATTAAGAATTCACAGGTTTTAGAAGATTGTCCAGGTGATCTTAATAAATG 1406
              |||  ||| || |||||  | |||||| | || || || || || ||||||| | | ||
Db        674 ACATGGCTACATAAAGAAAACGCAGGTTATCGATGACTGCCCTGGAGATCTTAGTCAGTG 733

Qy       1407 GTACCGTCACATTTCAAAAGGTGCTTGGCCTTTTTCAACTGGAGATCATGGATGGCCAAT 1466
              |||||| ||||| || |||||||| ||||| ||||| ||||  |||||||| ||||| ||
Db        734 GTACCGCCACATATCTAAAGGTGCATGGCCCTTTTCTACTGCTGATCATGGTTGGCCTAT 793

Qy       1467 TTCTGACTGCACAGCTGAAGGACTGAAAGCTGTTCTATTACTATCCAAAATTGCACCAGA 1526
              || || ||||| || |||||||| || || |   |||| ||||| || ||| | |||||
Db        794 ATCAGATTGCACTGCAGAAGGACTTAAGGCGGCATTATTGCTATCGAAGATTTCTCCAGA 853

Qy       1527 AATAGTTGGTGAGCCAATAGACGTGAAGCGATTATATGATTCTGTAAATGTCATTCTCTC 1586
              || ||||| ||  || | || || ||  || | |||||||||| |||    |  | ||
Db        854 TATTGTTGGCGAAGCAGTGGAAGTTAATAGACTGTATGATTCTGTCAATTGTTTGATGTC 913

Qy       1587 ACTACAGAATGAAGATGGTGGTTTTGCAACATATGAGCTTAAACGATCTTATAATTGGTT 1646
              |      ||||||  |||||| |||||||||||||| || | | | ||||||   ||| |
Db        914 ATACATGAATGATAATGGTGGATTTGCAACATATGAACTCACAAGGTCTTATGCCTGGCT 973

Qy       1647 GGAGATAATCAATCCTGCTGAAACTTTTGGTGACATCGTTATTGATTATCCTTATGTGGA 1706
              |||| | |||||||||||||| || ||||| ||||| || ||||||||||||||||||
Db        974 GGAGCTTATCAATCCTGCTGAGACCTTTGGGGACATTGTGATTGATTATCCTTATGTGGA 1033

Qy       1707 ATGTACATCAGCAGCGATTCAAGCTTTGGCATCATTTAGGAAATTATATCCTGGGCATCG 1766
              ||| || |||||||| ||||| ||  || || ||||||  ||  | || ||||| || ||
Db       1034 ATGCACTTCAGCAGCAATTCAGGCCCTGACAGCATTTAAAAAGCTCTACCCTGGACACCG 1093

Qy       1767 CCGAGAAGAAATACAACATTGTATCGATAAAGCCACTACCTTCATTGAAAAAATACAAGC 1826
              |      |||||| |  | |||||     |||||  ||| ||| |||||     || |||
Db       1094 CAAGAGTGAAATAGACAACTGTATAAGCAAAGCTGCTAGCTTTATTGAGGGTATTCAAAA 1153

Qy       1827 TTCAGATGGATCATGGTATGGTTCTTGGGGAGTTTGCTTCACTTACGGTGCTTGGTTTGG 1886
              |||||  |||||||||||||||||||  ||||| || || || ||  | ||||||||
Db       1154 AAGCGATGGTTCATGGTATGGTTCTTGGGCTGTTTGTTTTACCTATGGCACATGGTTTGG 1213

Qy       1887 GGTAAAAGGTCTGATTGCTGCTGGAAGGAGTTTCAGTAATTGCTCAAGCATCCGTAAAGC 1946
              ||||| ||  |  |||||||||| |||| ||||  ||  |  |    ||| | || ||
Db       1214 TGTAAAGGGATTAGTTGCTGCTGGTAGGACATTCAAAAACAGTCCTGCAATCAGAAAGGC 1273

Qy       1947 TTGTGAATTTCTGCTGTCCAAGCAGCTTCCTTCTGGTGGCTGGGGAGAGAGTTATCTGTC 2006
              ||||| ||| || |||| ||  |||||||||||| |||||||||| || ||| ||||
Db       1274 ATGTGACTTTTTGTTGTCAAAAGAGCTTCCTTCTGGAGGCTGGGGAGAAAGCTATTTGTC 1333

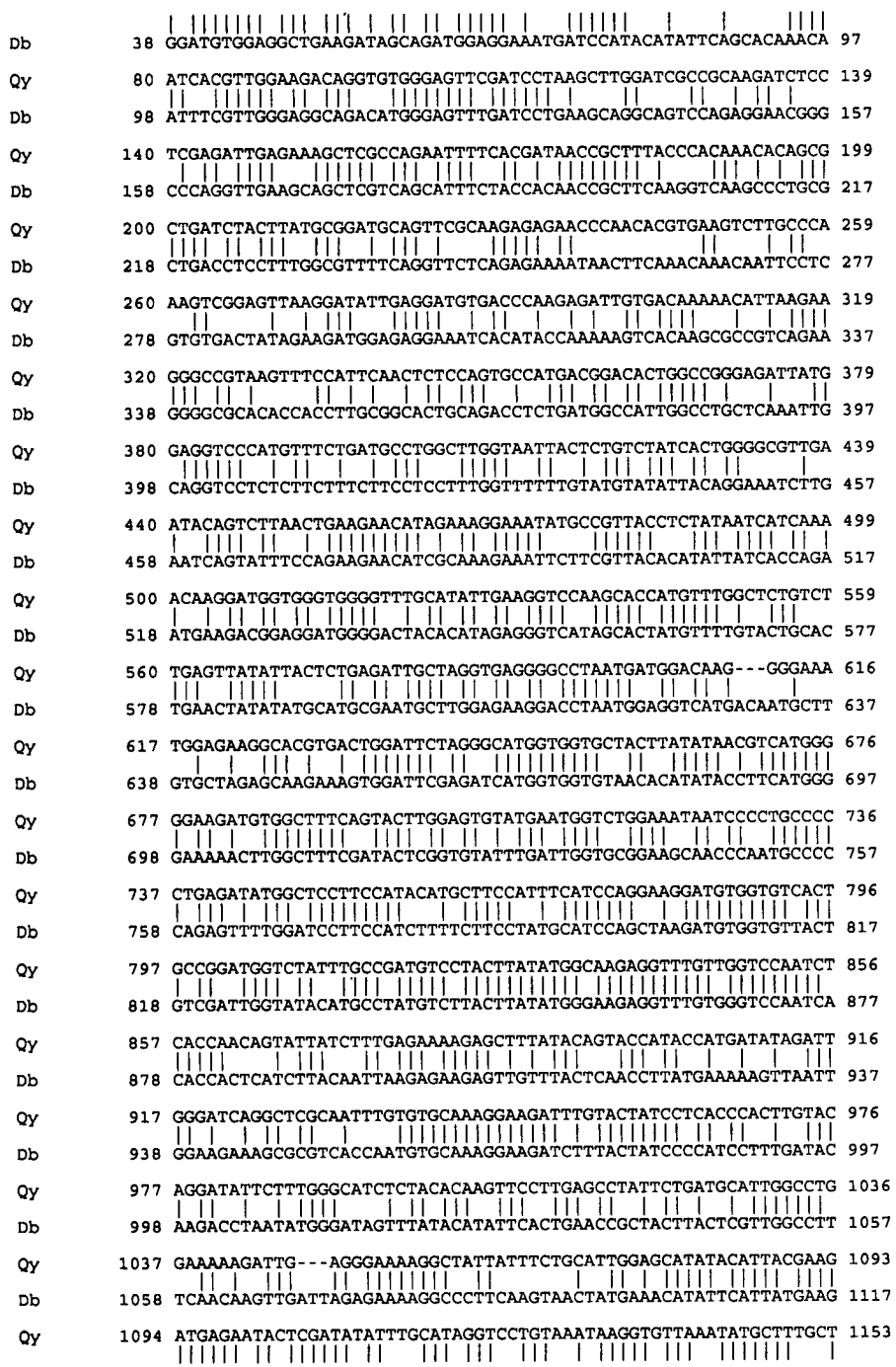
Qy       2007 CTGTCAAAACAAGGTGTATTCAAATCTGGAAGGCAACAGGTCTCATGTGGTCAACACTGG 2066
              |  ||| |  |||| ||| | ||||| |||| |  |||||| ||| ||||||||
Db       1334 ATCCCAAGATCAGGTTTATACCAATCTCGAAGGGAAGCGACCTCATGCGGTGAACACTGG 1393

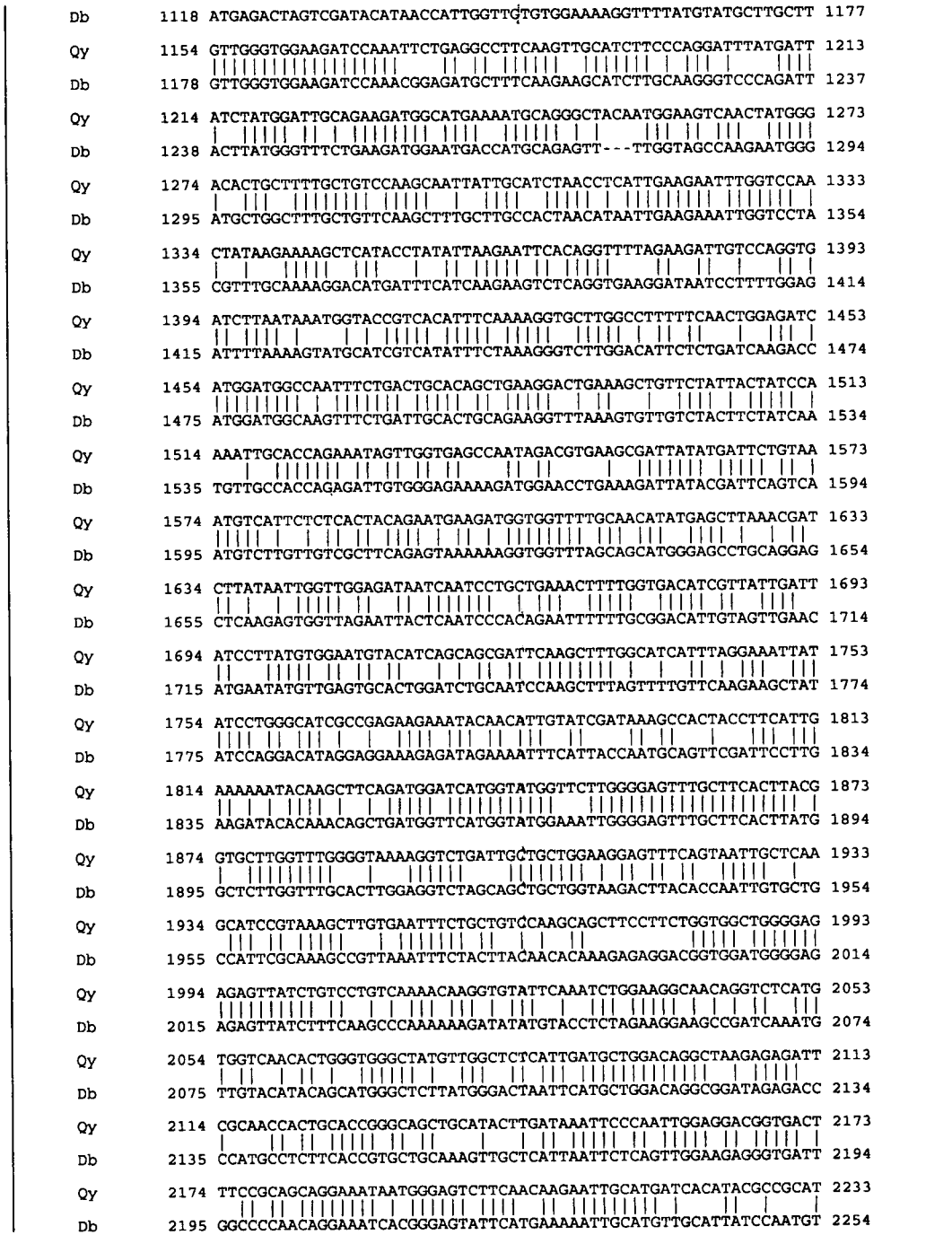
Qy       2067 GTGGGCTATGTTGGCTCTCATTGATGCTGGACAGGCTAAGAGAGATTCGCAACCACTGCA 2126
              ||||| ||| | || || ||||| |||||||| |||||||| |    ||  ||||
Db       1394 TTGGGCCATGCTAGCCCTAATCGATGCAGGGCAGGCTGAGAGAGATCCAATTCCTTTGCA 1453

Qy       2127 CCGGGCAGCTGCATACTTGATAAATTCCCAATTGGAGGACGGTGACTTTCCGCAGCAGGA 2186
              || |||||        ||||| || |  |||| ||| || ||||| ||||| ||||| ||
Db       1454 TCGAGCAGCGAAGGTTTTGATCAACTTACAATCGGAAGATGGTGAATTTCCCCAGCAAGA 1513

Qy       2187 AATAATGGGAGTCTTCAACAAGAATTGCATGATCACATACGCCGCATACAGAAACATATT 2246
              || || |||||||||||||| || ||||||||||  ||| |||  || |||||||| ||
Db       1514 GATCATTGGAGTCTTCAACAAAAACTGCATGATCAGCTACTCCGAGTATAGAAACATCTT 1573

Qy       2247 CCCCATTTGGGCGTTGGGAGAATACCAATCCCAAGTATTG 2286
              ||| ||||||||  | || || ||||    ||  || |||
Db       1574 CCCTATTTGGGCCCTTGGGGAGTACCGTCGCCGCGTCTTG 1613
```





```
Qy       560 TGAGTTATATTACTCTGAGATTGCTAGGTGAGGGGCCTAATGATGGACAAG---GGGAAA 616
Db       572 TGAACTATATATGCATGCGAATGCTTGGAGAAGGACCTAATGGAGGTCATGACAATGCTT 631

Qy       617 TGGAGAAGGCACGTGACTGGATTCTAGGGCATGGTGGTGCTACTTATATAACGTCATGGG 676
Db       632 GTGCTAGAGCAAGAAAGTGGATTCGAGATCATGGTGGTGTAACACATATACCTTCATGGG 691

Qy       677 GGAAGATGTGGCTTTCAGTACTTGGAGTGTATGAATGGTCTGGAAATAATCCCCTGCCCC 736
Db       692 GAAAAACTTGGCTTTCGATACTCGGTGTATTTGATTGGTGCGGAAGCAACCCAATGCCCC 751

Qy       737 CTGAGATATGGCTCCTTCCATACATGCTTCCATTTCATCCAGGAAGGATGTGGTGTCACT 796
Db       752 CAGAGTTTTGGATCCTTCCATCTTTTCTTCCTATGCATCCAGCTAAGATGTGGTGTTACT 811

Qy       797 GCCGGATGGTCTATTTGCCGATGTCCTACTTATATGGCAAGAGGTTTGTTGGTCCAATCT 856
Db       812 GTCGATTGGTATACATGCCTATGTCTTACTTATATGGGAAGAGGTTTGTGGGTCCAATCA 871

Qy       857 CACCAACAGTATTATCTTTGAGAAAAGAGCTTTATACAGTACCATACCATGATATAGATT 916
Db       872 CACCACTCATCTTACAATTAAGAGAAGAGTTGTTTACTCAACCTTATGAAAAAGTTAATT 931

Qy       917 GGGATCAGGCTCGCAATTTGTGTGCAAAGGAAGATTTGTACTATCCTCACCCACTTGTAC 976
Db       932 GGAAGAAAGCGCGTCACCAATGTGCAAAGGAAGATCTTTACTATCCCCATCCTTTGATAC 991

Qy       977 AGGATATTCTTTGGGCATCTCTACACAAGTTCCTTGAGCCTATTCTGATGCATTGGCCTG 1036
Db       992 AAGACCTAATATGGGATAGTTTATACATATTCACTGAACCGCTACTTACTCGTTGGCCTT 1051

Qy      1037 GAAAAAGATTG---AGGGAAAAGGCTATTATTTCTGCATTGGAGCATATACATTACGAAG 1093
Db      1052 TCAACAAGTTGATTAGAGAAAAGGCCCTTCAAGTAACTATGAAACATATTCATTATGAAG 1111

Qy      1094 ATGAGAATACTCGATATATTTGCATAGGTCCTGTAAATAAGGTGTTAAATATGCTTTGCT 1153
Db      1112 ATGAGACTAGTCGATACATAACCATTGGTTGTGTGGAAAAGGTTTTATGTATGCTTGCTT 1171

Qy      1154 GTTGGGTGGAAGATCCAAATTCTGAGGCCTTCAAGTTGCATCTTCCCAGGATTTATGATT 1213
Db      1172 GTTGGGTGGAAGATCCAAACGGAGATGCTTTCAAGAAGCATCTTGCAAGGGTCCCAGATT 1231

Qy      1214 ATCTATGGATTGCAGAAGATGGCATGAAAATGCAGGGCTACAATGGAAGTCAACTATGGG 1273
Db      1232 ACTTATGGGTTTCTGAAGATGGAATGACCATGCAGAGTT---TTGGTAGCCAAGAATGGG 1288

Qy      1274 ACACTGCTTTTGCTGTCCAAGCAATTATTGCATCTAACCTCATTGAAGAAATTTGGTCCAA 1333
Db      1289 ATGCTGGCTTTGCTGTTCAAGCTTTGCTTGCCACTAACATAATTGAAGAAATTGGTCCTA 1348

Qy      1334 CTATAAGAAAAGCTCATACCTATATTAAGAATTCACAGGTTTTAGAAGATTGTCCAGGTG 1393
Db      1349 CGTTTGCAAAAGGACATGATTTCATCAAGAAGTCTCAGGTGAAGGATAATCCTTTTGGAG 1408

Qy      1394 ATCTTAATAAATGGTACCGTCACATTTCAAAAGGTGCTTGGCCTTTTTCAACTGGAGATC 1453
Db      1409 ATTTTAAAAGTATGCATCGTCATATTTCTAAAGGGTCTTGGACATTCTCTGATCAAGACC 1468

Qy      1454 ATGGATGGCCAATTTCTGACTGCACAGCTGAAGGACTGAAAGCTGTTCTATTACTATCCA 1513
Db      1469 ATGGATGGCAAGTTTCTGATTGCACTGCAGAAGGTTTAAAGTGTTGTCTACTTCTATCAA 1528

Qy      1514 AAATTGCACCAGAAATAGTTGGTGAGCCAATAGACGTGAAGCGATTATATGATTCTGTAA 1573
Db      1529 TGTTGCCACCAGAGATTGTGGGAGAAAAGATGGAACCTGAAAGATTATACGATTCAGTCA 1588

Qy      1574 ATGTCATTCTCTCACTACAGAATGAAGATGGTGGTTTTGCAACATATGAGCTTAAACGAT 1633
Db      1589 ATGTCTTGTTGTCGCTTCAGAGTAAAAAAGGTGGTTTAGCAGCATGGGAGCCTGCAGGAG 1648

Qy      1634 CTTATAATTGGTTGGAGATAATCAATCCTGCTGAAACTTTTGGTGACATCGTTATTGATT 1693
Db      1649 CTCAAGAGTGGTTAGAATTACTCAATCCCACAGAATTTTTTGCGGACATTGTAGTTGAAC 1708

Qy      1694 ATCCTTATGTGGAATGTACATCAGCAGCGATTCAAGCTTTGGCATCATTTAGGAAATTAT 1753
Db      1709 ATGAATATGTTGAGTGCACTGGATCTGCAATCCAAGCTTTAGTTTTGTTCAAGAAGCTAT 1768

Qy      1754 ATCCTGGGCATCGCCGAGAAGAAATACAACATTGTATCGATAAAGCCACTACCTTCATTG 1813
Db      1769 ATCCAGGACATAGGAAGAAAGAGATAGAAAATTTCATTACCAATGCAGTTCGATTCCTTG 1828

Qy      1814 AAAAAATACAAGCTTCAGATGGATCATGGTATGGTTCTTGGGGAGTTTGCTTCACTTACG 1873
Db      1829 AAGATACACAAACAGCTGATGGTTCATGGTATGGAAATTGGGGAGTTTGCTTCACTTATG 1888

Qy      1874 GTGCTTGGTTTGGGGTAAAAGGTCTGATTGCTGCTGGAAGGAGTTTCAGTAATTGCTCAA 1933
Db      1889 GCTCTTGGTTTGCACTTGGAGGTCTAGCAGCTGCTGGTAAGACTTACACCAATTGTGCTG 1948

Qy      1934 GCATCCGTAAAGCTTGTGAATTTCTGCTGTCCAAGCAGCTTCCTTCTGGTGGCTGGGGAG 1993
Db      1949 CCATTCGCAAAGCCGTTAAATTTCTACTTACAACACAAAGAGAGGACGGTGGATGGGGAG 2008

Qy      1994 AGAGTTATCTGTCCTGTCAAAACAAGGTGTATTCAAATCTGGAAGGCAACAGGTCTCATG 2053
Db      2009 AGAGTTATCTTTCAAGCCCAAAAAAGATATATGTACCTCTAGAAGGAAGCCGATCAAATG 2068

Qy      2054 TGGTCAACACTGGGTGGGCTATGTTGGCTCTCATTGATGCTGGACAGGCTAAGAGAGATT 2113
Db      2069 TTGTACATACAGCATGGGCTCTTATGGGACTAATTCATGCTGGACAGGCGGATAGAGACC 2128

Qy      2114 CGCAACCACTGCACCGGGCAGCTGCATACTTGATAAATTCCCAATTGGAGGACGGTGACT 2173
Db      2129 CCATGCCTCTTCACCGTGCTGCAAAGTTGCTCATTAATTCTCAGTTGGAAGAGGGTGATT 2188

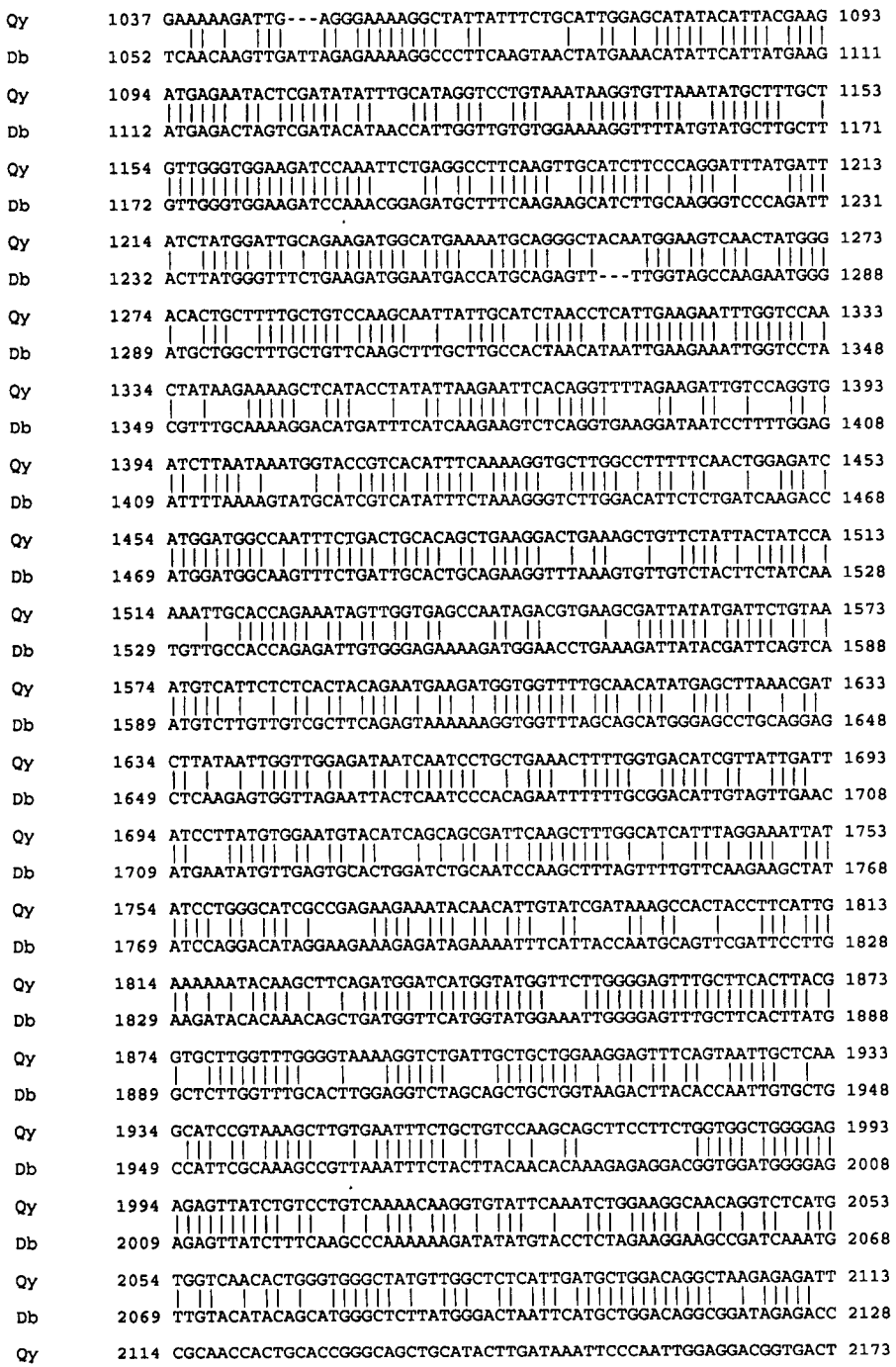
Qy      2174 TTCCGCAGCAGGAAATAATGGGAGTCTTCAACAAGAATTGCATGATCACATACGCCGCAT 2233
Db      2189 GGCCCCAACAGGAAATCACGGGAGTATTCATGAAAAATTGCATGTTGCATTATCCAATGT 2248

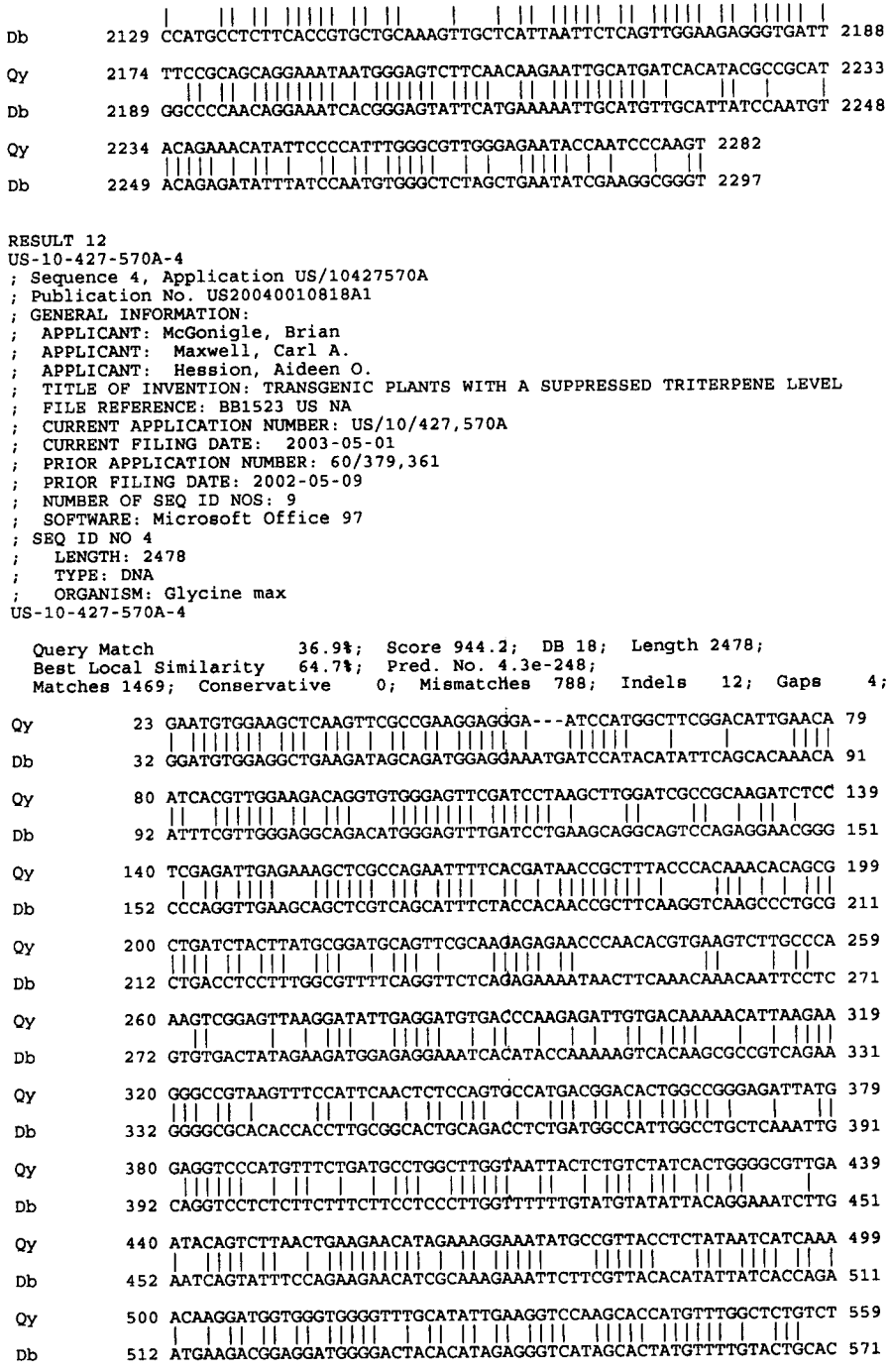
Qy      2234 ACAGAAACATATTCCCCATTTGGGCGTTGGGAGAATACCAATCCCAAGT 2282
Db      2249 ACAGAGATATTTATCCAATGTGGGCTCTAGCTGAATATCGAAGGCGGGT 2297
```

RESULT 13
US-10-424-599-35539
; Sequence 35539, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 35539
; LENGTH: 2669
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_132095C.1
US-10-424-599-35539

Query Match 36.9%; Score 944.2; DB 19; Length 2669;
Best Local Similarity 64.7%; Pred. No. 4.5e-248;
Matches 1469; Conservative 0; Mismatches 788; Indels 12; Gaps 4;

```
Qy        23 GAATGTGGAAGCTCAAGTTCGCCGAAGGAGGGA---ATCCATGGCTTCGGACATTGAACA 79
```





RESULT 12
US-10-427-570A-4
; Sequence 4, Application US/10427570A
; Publication No. US20040010818A1
; GENERAL INFORMATION:
; APPLICANT: McGonigle, Brian
; APPLICANT: Maxwell, Carl A.
; APPLICANT: Hession, Aideen O.
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH A SUPPRESSED TRITERPENE LEVEL
; FILE REFERENCE: BB1523 US NA
; CURRENT APPLICATION NUMBER: US/10/427,570A
; CURRENT FILING DATE: 2003-05-01
; PRIOR APPLICATION NUMBER: 60/379,361
; PRIOR FILING DATE: 2002-05-09
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 4
; LENGTH: 2478
; TYPE: DNA
; ORGANISM: Glycine max
US-10-427-570A-4

Query Match 36.9%; Score 944.2; DB 18; Length 2478;
Best Local Similarity 64.7%; Pred. No. 4.3e-248;
Matches 1469; Conservative 0; Mismatches 788; Indels 12; Gaps 4;

```
Qy       1609 TTTGCAACATATGAGCTTAAACGATCTTATAATTGGTTGGAGATAATCAATCCTGCTGAA 1668
Db       1726 TTTGCGACTTATGAACTCACAAGATCCTATGCCTGGCTGGAGCTGATCAATCCTGCTGAG 1785

Qy       1669 ACTTTTGGTGACATCGTTATTGATTATCCTTATGTGGAATGTACATCAGCAGCGATTCAA 1728
Db       1786 ACCTTTGGGGATATTGTGATTGATTACCC--------------------CAGCAATCCAG 1825

Qy       1729 GCTTTGGCATCATTTAGGAAATTATATCCTGGGCATCGCCGAGAAGAAATACAACATTGT 1788
Db       1826 GCCCTGACATCATTCAGAAAACTATACCCTGGGCACCGCAGGAAAGAGGTGGACAGCTGT 1885

Qy       1789 ATCGATAAAGCCACTACCTTCATTGAAAAAATACAAGCTTCAGATGGATCATGGTATGGT 1848
Db       1886 ATCAGCAAAGCTGCTAATTTCATTGAGAGCATTCAGAAAAGTGACGGTTCATGGTATGGC 1945

Qy       1849 TCTTGGGGAGTTTGCTTCACTTACGGTGCTTGGTTTGGGGTAAAAGGTCTGATTGCTGCT 1908
Db       1946 TCTTGGGCTGTTTGTTTCACGTACGGCACTTGGTTTGGTGTGAAGGGACTAATTGCTGCT 2005

Qy       1909 GGAAGGAGTTTCAGTAATTGCTCAAGCATCCGTAAAGCTTGTGAATTTCTGCTGTCCAAG 1968
Db       2006 GGTAGAACATTTGAGAGCAGTCCTGCAATCAGAAAGGCATGCGGCTTTCTGTTATCAAAA 2065

Qy       1969 CAGCTTCCTTCTGGTGGCTGGGGAGAGAGTTATCTGTCCTGTCAAAACAAGGTGTATTCA 2028
Db       2066 GAGCTTCCTTCCGGTGGCTGGGGAGAAAGCTATTTATCGTCTCAAGACCAGGTTTACACC 2125

Qy       2029 AATCTGGAAGGCAACAGGTCTCATGTGGTCAACACTGGGTGGGCTATGTTGGCTCTCATT 2088
Db       2126 AATCTCAAAGGCAACCGAGCTCATGCAGCGAACACTGGTTGGGCCATGCTGGCGCTGATT 2185

Qy       2089 GATGCTGGACAGGCTAAGAGAGATTCGCAACCACTGCACCGGGCAGCTGCATACTTGATA 2148
Db       2186 GATGCAGGACAGGCCCAGAGAGATCCCGCGCCTCTACACCGAGCAGCAAGGGTTTTGATA 2245

Qy       2149 AATTCCCAATTGGAGGACGGTGACTTTCCGCAGCAGGAAATAATGGGAGTCTTCAACAAG 2208
Db       2246 AACTTGCAGTCAGAGGACGGAGAATTTCCCCAGCAAGAGATCATAGGAGTGTTCAACAAG 2305

Qy       2209 AATTGCATGATCACATACGCCGCATACAGAAACATATTCCCCATTTGGGCGTTGGGAGAA 2268
Db       2306 AACTGCATGATCAGTTACTCCCAGTACAGGAACACCTTCCCGATCTGGGCTCTGGGAGAG 2365

Qy       2269 TACCAATCCCAAGTATTGCAATCTCG 2294
Db       2366 TACCGGTGCCGAGTCCTGGGAGCTGG 2391
```

```
RESULT 11
US-10-221-074-3
; Sequence 3, Application US/10221074
; Publication No. US20030208791A1
; GENERAL INFORMATION:
;  APPLICANT: Bryan, Gregory T.
;  APPLICANT:  Hwang, Der-Chyan
;  APPLICANT:  Maxwell, Carl A.
;  APPLICANT:  McGonigle, Brian
;  APPLICANT:  Potter, Susan M.
;  TITLE OF INVENTION: ENZYMES INVOLVED IN TRITERPENE SYNTHESIS
;  FILE REFERENCE: BB1438 PCT
;  CURRENT APPLICATION NUMBER: US/10/221,074
;  CURRENT FILING DATE:  2003-01-08
;  PRIOR APPLICATION NUMBER: 60/188,054
;  PRIOR FILING DATE: 2000-03-09
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 3
;   LENGTH: 2478
;   TYPE: DNA
;   ORGANISM: Glycine max
US-10-221-074-3
```

Query Match 36.9%; Score 944.2; DB 18; Length 2478;
Best Local Similarity 64.7%; Pred. No. 4.3e-248;
Matches 1469; Conservative 0; Mismatches 788; Indels 12; Gaps 4;

```
Qy         23 GAATGTGGAAGCTCAAGTTCGCCGAAGGAGGGA---ATCCATGGCTTCGGACATTGAACA 79
Db         32 GGATGTGGAGGCTGAAGATAGCAGATGGAGGAAATGATCCATACATATTCAGCACAAACA 91

Qy         80 ATCACGTTGGAAGACAGGTGTGGGAGTTCGATCCTAAGCTTGGATCGCCGCAAGATCTCC 139
Db         92 ATTTCGTTGGGAGGCAGACATGGGAGTTTGATCCTGAAGCAGGCAGTCCAGAGGAACGGG 151

Qy        140 TCGAGATTGAGAAAGCTCGCCAGAATTTTCACGATAACCGCTTTACCCACAAACACAGCG 199
Db        152 CCCAGGTTGAAGCAGCTCGTCAGCATTTCTACCACAACCGCTTCAAGGTCAAGCCCTGCG 211

Qy        200 CTGATCTACTTATGCGGATGCAGTTCGCAAGAGAGAACCCAACACGTGAAGTCTTGCCCA 259
Db        212 CTGACCTCCTTTGGCGTTTTCAGGTTCTCAGAGAAAATAACTTCAAACAAACAATTCCTC 271

Qy        260 AAGTCGGAGTTAAGGATATTGAGGATGTGACCCAAGAGATTGTGACAAAAACATTAAGAA 319
Db        272 GTGTGACTATAGAAGATGGAGAGGAAATCACATACCAAAAAGTCACAAGCGCCGTCAGAA 331

Qy        320 GGGCCGTAAGTTTCCATTCAACTCTCCAGTGCCATGACGGACACTGGCCGGGAGATTATG 379
Db        332 GGGGCGCACACCACCTTGCGGCACTGCAGACCTCTGATGGCCATTGGCCTGCTCAAATTG 391

Qy        380 GAGGTCCCATGTTTCTGATGCCTGGCTTGGTAATTACTCTGTCTATCACTGGGGCGTTGA 439
Db        392 CAGGTCCTCTCTTCTTTCTTCCTCCCTTGGTTTTTTGTATGTATATTACAGGAAATCTTG 451

Qy        440 ATACAGTCTTAACTGAAGAACATAGAAAGGAAATATGCCGTTACCTCTATAATCATCAAA 499
Db        452 AATCAGTATTTCCAGAAGAACATCGCAAAGAAATTCTTCGTTACACATATTATCACCAGA 511

Qy        500 ACAAGGATGGTGGGTGGGGTTTGCATATTGAAGGTCCAAGCACCATGTTTGGCTCTGTCT 559
Db        512 ATGAAGACGGAGGATGGGGACTACACATAGAGGGTCATAGCACTATGTTTTGTACTGCAC 571

Qy        560 TGAGTTATATTACTCTGAGATTGCTAGGTGAGGGGCCTAATGATGGACAAG---GGGAAA 616
Db        572 TGAACTATATATGCATGCGAATGCTTGGAGAAGGACCTAATGGAGGTCATGACAATGCTT 631

Qy        617 TGGAGAAGGCACGTGACTGGATTCTAGGGCATGGTGGTGCTACTTATATAACGTCATGGG 676
Db        632 GTGCTAGAGCAAGAAAGTGGATTCGAGATCATGGTGGTGTAACACATATACCTTCATGGG 691

Qy        677 GGAAGATGTGGCTTTCAGTACTTGGAGTGTATGAATGGTCTGGAAATAATCCCCTGCCCC 736
Db        692 GAAAAACTTGGCTTTCGATACTCGGTGTATTTGATTGGTGCGGAAGCAACCCAATGCCCC 751

Qy        737 CTGAGATATGGCTCCTTCCATACATGCTTCCATTTCATCCAGGAAGGATGTGGTGTCACT 796
Db        752 CAGAGTTTTGGATCCTTCCATCTTTTCTTCCTATGCATCCAGCTAAGATGTGGTGTTACT 811

Qy        797 GCCGGATGGTCTATTTGCCGATGTCCTACTTATATGGCAAGAGGTTTGTTGGTCCAATCT 856
Db        812 GTCGATTGGTATACATGCCTATGTCTTACTTATATGGGAAGAGGTTTGTGGGTCCAATCA 871

Qy        857 CACCAACAGTATTATCTTTGAGAAAAGAGCTTTATACAGTACCATACCATGATATAGATT 916
Db        872 CACCACTCATCTTACAATTAAGAGAAGAGTTGTTTACTCAACCTTATGAAAAAGTTAATT 931

Qy        917 GGGATCAGGCTCGCAATTTGTGTGCAAAGGAAGATTTGTACTATCCTCACCCACTTGTAC 976
Db        932 GGAAGAAAGCGCGTCACCAATGTGCAAAGGAAGATCTTTACTATCCCCATCCTTTGATAC 991

Qy        977 AGGATATTCTTTGGGCATCTCTACACAAGTTCCTTGAGCCTATTCTGATGCATTGGCCTG 1036
Db        992 AAGACCTAATATGGGATAGTTTATACATATTCACTGAACCGCTACTTACTCGTTGGCCTT 1051
```

```
Qy       2133 AGCTGCATACTTGATAAATTCCCAATTGGAGGACGGTGACTTTCCGCAGCAGGAAATAAT 2192
Db       2187 AGCGAAGGTTTTGATCAACTTACAATCGGAAGATGGTGAATTTCCCCAGCAAGAGATCAT 2246

Qy       2193 GGGAGTCTTCAACAAGAATTGCATGATCACATACGCCGCATACAGAAACATATTCCCCAT 2252
Db       2247 TGGAGTCTTCAACAAAAACTGCATGATCAGCTACTCCGAGTATAGAAACATCTTCCCTAT 2306

Qy       2253 TTGGGCGTTGGGAGAATACCAATCCCAAGTATTG 2286
Db       2307 TTGGGCCCTTGGGGAGTACCGTCGCCGCGTCTTG 2340
```

RESULT 10
US-10-425-114-24593
; Sequence 24593, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 24593
; LENGTH: 2653
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3632-042-C4_FLI
US-10-425-114-24593

Query Match 44.7%; Score 1145.6; DB 19; Length 2653;
Best Local Similarity 70.3%; Pred. No. 1.8e-303;
Matches 1578; Conservative 0; Mismatches 639; Indels 29; Gaps 2;

```
Qy         49 GGAGGGAATCCATGGCTTCGGACATTGAACAATCACGTTGGAAGACAGGTGTGGGAGTTC 108
Db        175 GGCGGGGACCCCTGGCTCCGCACGAAGAACGCCCACGTCGGCCGCCAGGTCTGGGAGTTC 234

Qy        109 GATCCTAAGCTTGGATCGCCGCAAGATCTCCTCGAGATTGAGAAAGCTCGCCAGAATTTT 168
Db        235 GACGCC---------TCCGCCGACCCCGACCCCGCCGTCGACGCCGCGCGCCGCGCCTTC 285

Qy        169 CACGATAACCGCTTTACCCACAAACACAGCGCTGATCTACTTATGCGGATGCAGTTCGCA 228
Db        286 GCCGGCAGCAGGGGCCACCTCAAGCACAGCGCCGACCTCCTCATGCGCATCCAGTTCGCC 345

Qy        229 AGAGAGAACCCAACACGTGAAGTCTTGCCCAAAGTCGGAGTTAAGGATATTGAGGATGTG 288
Db        346 AAAGAAAATCCTCTTGAGCTTGACCTTCCAGCGATCAAGCTCGGTGAGCACGAAGACGTA 405

Qy        289 ACCCAAGAGATTGTGACAAAAACATTAAGAAGGGCCGTAAGTTTCCATTCAACTCTCCAG 348
Db        406 ACCGAGGAAGCCGTGTCTACTACTTTGAGAAGGGCAATCAGCCGTCTTTCTACTCTCCAG 465

Qy        349 TGCCATGACGGACACTGGCCGGGAGATTATGGAGGTCCCATGTTTCTGATGCCTGGCTTG 408
Db        466 GCACATGATGGACACTGGCCTGGGGATTATGGTGGCCCTATGTTCCTTATGCCAGGCTTG 525

Qy        409 GTAATTACTCTGTCTATCACTGGGGCGTTGAATACAGTCTTAACTGAAGAACATAGAAAG 468
Db        526 ATCATAACATTGTATGTGACTGGAGCACTAAACACTGTCTTGTCATCAGAACACCAGAAG 585

Qy        469 GAAATATGCCGTTACCTCTATAATCATCAAAACAAGGATGGTGGGTGGGGTTTGCATATT 528
Db        586 GAGATCCGTCGGTATCTTTATAATCATCAGAATGAAGATGGTGGCTGGGGCTTGCACATT 645

Qy        529 GAAGGTCCAAGCACCATGTTTGGCTCTGTCTTGAGTTATATTACTCTGAGATTGCTAGGT 588
Db        646 GAGGGTCCAAGCACCATGTTTGGCTCAGCACTGACCTACGTTACTTTGAGATTGCTTGGA 705

Qy        589 GAGGGGCCTAATGATGGACAAGGGGAAATGGAGAAGGCACGTGACTGGATTCTAGGGCAT 648
Db        706 GAGGGACCAGATAGTGGAGATGGAGCAATGGAGAAAGGCCGAAACTGGATCTTAGACCAT 765

Qy        649 GGTGGTGCTACTTATATAACGTCATGGGGGAAGATGTGGCTTTCAGTACTTGGAGTGTAT 708
Db        766 GGCGGAGCAACATATATAACATCCTGGGGGAAGTTTTGGCTCTCGGTACTTGGTGTATTT 825

Qy        709 GAATGGTCTGGAAATAATCCCCTGCCCCCTGAGATATGGCTCCTTCCATACATGCTTCCA 768
Db        826 GAATGGTCTGGTAACAACCCGGTACCACCAGAAGTATGGCTACTGCCATATCTCCTGCCA 885

Qy        769 TTTCATCCAGGAAGGATGTGGTGTCACTGCCGGATGGTCTATTTGCCGATGTCCTACTTA 828
Db        886 TTTCACCCAGGGAGGATGTGGTGTCATTGTCGAATGGTGTATTTGCCAATGTGCTACATC 945

Qy        829 TATGGCAAGAGGTTTGTTGGTCCAATCTCACCAACAGTATTATCTTTGAGAAAAGAGCTT 888
Db        946 TATGGGAAGAGGTTTGTTGGTCGAATCACACCACTCGTGTTGGAATTAAGAAAGGAACTT 1005

Qy        889 TATACAGTACCATACCATGATATAGATTGGGATCAGGCTCGCAATTTGTGTGCAAAGGAA 948
Db       1006 TTCAAAGACCCTTACAGCAAGATTGATTGGGACAAGGCCCGCAATCTATGTGCTAAGGAA 1065

Qy        949 GATTTGTACTATCCTCACCCACTTGTACAGGATATTCTTTGGGCATCTCTACACAAGTTC 1008
Db       1066 GATCTGTACTATCCACATCCATTTGTTCAAGATGTCTTATGGGCCACTCTCCATAAATTT 1125

Qy       1009 CTTGAGCCTATTCTGATGCATTGGCCTGGAAAAAGATTGAGGGAAAAGGCTATTATTTCT 1068
Db       1126 GTTGAACCAGTTATGATGAGTTGGCCTGGCAGTAAATTGAGGGAGAAAGCTCTGGAGACA 1185

Qy       1069 GCATTGGAGCATATACATTACGAAGATGAGAATACTCGATATATTTGCATAGGTCCTGTA 1128
Db       1186 GCCATGCAACATGTTCACTATGAGGATGAGAACACTCGGTACATTTGCATTGGTCCTGTA 1245

Qy       1129 AATAAGGTGTTAAATATGCTTTGCTGTTGGGTGGAAGATCCAAATTCTGAGGCCTTCAAG 1188
Db       1246 AACAAGGTATTAAATATGCTTGCTTGCTGGATTGAAGATCCAAACTCAGAGGCCTTCAAA 1305

Qy       1189 TTGCATCTTCCCAGGATTTATGATTATCTATGGATTGCAGAAGATGGCATGAAAATGCAG 1248
Db       1306 CTCCACATCCCACGAGTCTATGACTACTTGTGGCTTGCTGAAGATGGCATGAAGATGCAG 1365

Qy       1249 GGCTACAATGGAAGTCAACTATGGGACACTGCTTTTGCTGTCCAAGCAATTATTGCATCT 1308
Db       1366 GGTTATAATGGTAGCCAACTTTGGGATACAGCTTTCATAGTTCAAGCCATTGTGGCTACC 1425

Qy       1309 AACCTCATTGAAGAATTTGGTCCAACTATAAGAAAAGCTCATACCTATATTAAGAATTCA 1368
Db       1426 AACCTTACCGAAGAGTTTGGTCCCACCCTTAAACTAGCACACAACTATATCAAGAAATCA 1485

Qy       1369 CAGGTTTTAGAAGATTGTCCAGGTGATCTTAATAAATGGTACCGTCACATTTCAAAAGGT 1428
Db       1486 CAGGTTCTTGACGACTGCCCTGGTGATCTGAATGACTGGTACCGCCACACATCTAAAGGC 1545

Qy       1429 GCTTGGCCTTTTTCAACTGGAGATCATGGATGGCCAATTTCTGACTGCACAGCTGAAGGA 1488
Db       1546 GCATGGCCATTCTCCACCGCGGATCATGGTTGGCCTATATCTGATTGCACTGCTGAAGGA 1605

Qy       1489 CTGAAAGCTGTTCTATTACTATCCAAAATTGCACCAGAAATAGTTGGTGAGCCAATAGAC 1548
Db       1606 CTAAAGGCTTCATTATTGTTATCAAAGATTTCTCCTGAAATTGTTGGTGGACCGATAGAA 1665

Qy       1549 GTGAAGCGATTATATGATTCTGTAAATGTCATTCTCTCACTACAGAATGAAGATGGTGGT 1608
Db       1666 GCTAATAGATTTTATGATGCTGTCAGTTGTCTGATGTCTTATATGAATGATAATGGTGGT 1725
```

```
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 71159
;   LENGTH: 2355
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_7165C.1
US-10-437-963-71159

  Query Match             45.4%;  Score 1162;  DB 20;  Length 2355;
  Best Local Similarity   72.5%;  Pred. No. 5.1e-308;
  Matches 1504;  Conservative    0;  Mismatches  570;  Indels    0;  Gaps    0;

Qy        213 GCGGATGCAGTTCGCAAGAGAGAACCCAACACGTGAAGTCTTGCCCAAAGTCGGAGTTAA 272
              | | |  ||||| || | ||  || ||          | | | ||     ||    | |
Db        267 GAGAACTCAGTTTGCTAAAGCAAATCCTCTTAAGCTGGACATTCCAGCTATCAAGCTAGA 326

Qy        273 GGATATTGAGGATGTGACCCAAGAGATTGTGACAAAAACATTAAGAAGGGCCGTAAGTTT 332
              |||   ||| | ||| ||     ||   |||    |   ||| ||||||  |   |
Db        327 GGAGCATGAAGCTGTTACAGGCGAAGCTGTATTATCCTCTCTAAAAAGGGCAATTGCTCG 386

Qy        333 CCATTCAACTCTCCAGTGCCATGACGGACACTGGCCGGGAGATTATGGAGGTCCCATGTT 392
              ||||  |||  | ||    ||||| |||||||||||| || ||||||||| ||||||||||
Db        387 TTATTCTACTTTTCAAGCACATGATGGACACTGGCCTGGGGATTATGGTGGTCCCATGTT 446

Qy        393 TCTGATGCCTGGCTTGGTAATTACTCTGTCTATCACTGGGGCGTTGAATACAGTCTTAAC 452
              || ||||| |||||| | || ||  | | | |  || ||| |||| || |  ||  |
Db        447 CCTTATGCCAGGCTTGATTATAACATTATATGTGAGCGGAGCGCTGAACACTGCTTTGTC 506

Qy        453 TGAAGAACATAGAAAGGAAATATGCCGTTACCTCTATAATCATCAAAACAAGGATGGTGG 512
                  ||||||   ||||| ||| | || || ||||| |||||||| ||  | ||||| ||
Db        507 ATCTGAACATCAGAAGGAGATACGGCGGTATCTCTACAATCATCAGAATGAAGATGGAGG 566

Qy        513 GTGGGGTTTGCATATTGAAGGTCCAAGCACCATGTTTGGCTCTGTCTTGAGTTATATTAC 572
              ||||||||||| ||||| || |  |||||||||||||||| |  ||||  ||| || | |
Db        567 ATGGGGTTTGCACATTGAGGGCCACAGCACCATGTTTGGCTCAGCGTTGACATATGTTTC 626

Qy        573 TCTGAGATTGCTAGGTGAGGGGCCTAATGATGGACAAGGGGAAATGGAGAAGGCACGTGA 632
              | |||||||||| || ||||||||  ||    ||  | || |  |||||||| | |||  |
Db        627 TTTGAGATTGCTTGGGGAGGGGCCAGATAGCGGCGATGGAGCTATGGAGAAAGGACGAAA 686

Qy        633 CTGGATTCTAGGGCATGGTGGTGCTACTTATATAACGTCATGGGGGAAGATGTGGCTTTC 692
              ||||||  |||  ||||| || || |||||||| || |||||||| |||  | |||||||
Db        687 ATGGATTTTAGACCATGGGGGAGCAACTTATATTACATCATGGGGCAAGTTTTGGCTTTC 746

Qy        693 AGTACTTGGAGTGTATGAATGGTCTGGAAATAATCCCCTGCCCCCTGAGATATGGCTCCT 752
              || ||||| || | ||| |||||||| || || ||  |||| || || |||||| |  |
Db        747 GGTGCTTGGTGTATTTGACTGGTCTGGCAACAACCCAGTGCCACCAGAAATATGGTTGTT 806

Qy        753 TCCATACATGCTTCCATTTCATCCAGGAAGGATGTGGTGTCACTGCCGGATGGTCTATTT 812
              |||||  | || ||  |||||||||||  | |||||||||||| |||||||||| |||||
Db        807 GCCATATTTCCTGCCGATTCATCCAGGGCGAATGTGGTGTCATTGCCGGATGGTTTATTT 866

Qy        813 GCCGATGTCCTACTTATATGGCAAGAGGTTTGTTGGTCCAATCTCACCAACAGTATTATC 872
              ||| ||||  ||| | ||||| |||||||||| || ||| |  ||||||   ||||
Db        867 GCCTATGTGTTACATTTATGGAAAGAGGTTTGTGGGCCCAGTTACACCAATTATATTGGA 926

Qy        873 TTTGAGAAAAGAGCTTTATACAGTACCATACCATGATATAGATTGGGATCAGGCTCGCAA 932
              || ||||| || || ||   |||||| ||| ||||  | ||||||| ||||||||||
Db        927 ATTAAGAAAGGAACTCTACGAAGTACCCTACAATGAAGTTGATTGGGACAAGGCTCGCAA 986

Qy        933 TTTGTGTGCAAAGGAAGATTTGTACTATCCTCACCCACTTGTACAGGATATTCTTTGGGC 992
              | | ||||| ||||||||| |||||||||| || ||| | || ||||||  |  | |||||
Db        987 TCTATGTGCTAAGGAAGATCTGTACTATCCACATCCATTCGTGCAGGATGTATTATGGGC 1046

Qy        993 ATCTCTACACAAGTTCCTTGAGCCTATTCTGATGCATTGGCCTGGAAAAAGATTGAGGGA 1052
                  |||| ||||| ||  |||| ||   | || ||| ||||||||| || | |||||||||

Db       1047 CACTCTCCACAAATTTGTTGAACCAGCTATGTTGCGTTGGCCTGGGAACAAATTGAGGGA 1106

Qy       1053 AAAGGCTATTATTTCTGCATTGGAGCATATACATTACGAAGATGAGAATACTCGATATAT 1112
              || ||| |      |||   || ||||||| ||||| |||||||||| || |||||||||
Db       1107 GAAAGCTTTGGACACTGTCATGCAGCATATTCATTATGAAGATGAGAACACCCGATATAT 1166

Qy       1113 TTGCATAGGTCCTGTAAATAAGGTGTTAAATATGCTTTGCTGTTGGGTGGAAGATCCAAA 1172
              |||||| ||||| ||||| ||||| |||||||||||    || ||| | ||||||||||
Db       1167 TTGCATTGGTCCAGTAAACAAGGTATTAAATATGCTTGCTTGCTGGATTGAAGATCCAAA 1226

Qy       1173 TTCTGAGGCCTTCAAGTTGCATCTTCCCAGGATTTATGATTATCTATGGATTGCAGAAGA 1232
              || ||||| ||||| | ||  |||| ||  |  | ||||| ||||||||||||||||||
Db       1227 CTCAGAGGCATTCAAACTCCACATTCCAAGAGTCCACGATTACCTATGGATTGCAGAAGA 1286

Qy       1233 TGGCATGAAAATGCAGGGCTACAATGGAAGTCAACTATGGGACACTGCTTTTGCTGTCCA 1292
              |||||||||||||||||| |||||||| || || |||||||| |||||  | || ||
Db       1287 TGGCATGAAAATGCAGGGTTATAATGGAAGCCAGCTGTGGGACACAGCTTTCACAGTTCA 1346

Qy       1293 AGCAATTATTGCATCTAACCTCATTGAAGAATTTGGTCCAACTATAAGAAAAGCTCATAC 1352
              ||| ||  | ||  ||  |||||||||||||||||||||| ||| | | |  ||| |||
Db       1347 AGCTATAGTGGCTACTGGCCTCATTGAAGAATTTGGTCCTACTCTTAAACTAGCACATGG 1406

Qy       1353 CTATATTAAGAATTCACAGGTTTTAGAAGATTGTCCAGGTGATCTTAATAAATGGTACCG 1412
              ||| || |||||  | |||||| | || || || || || ||||||| | | ||||||||
Db       1407 CTACATAAAGAAAACGCAGGTTATCGATGACTGCCCTGGAGATCTTAGTCAGTGGTACCG 1466

Qy       1413 TCACATTTCAAAAGGTGCTTGGCCTTTTTCAACTGGAGATCATGGATGGCCAATTTCTGA 1472
              ||||| || |||||||| ||||| |||| |||||||| |||| || || ||
Db       1467 CCACATATCTAAAGGTGCATGGCCCTTTTCTACTGCTGATCATGGTTGGCCTATATCAGA 1526

Qy       1473 CTGCACAGCTGAAGGACTGAAAGCTGTTCTATTACTATCCAAAATTGCACCAGAAATAGT 1532
              ||||| || |||||||| || || |   |||| ||||| || ||| | ||||| || ||
Db       1527 TTGCACTGCAGAAGGACTTAAGGCGGCATTATTGCTATCGAAGATTTCTCCAGATATTGT 1586

Qy       1533 TGGTGAGCCAATAGACGTGAAGCGATTATATGATTCTGTAAATGTCATTCTCTCACTACA 1592
              ||| ||  || | || || ||  || | ||||||||||| |||     | | |||
Db       1587 TGGCGAAGCAGTGGAAGTTAATAGACTGTATGATTCTGTCAATTGTTTGATGTCATACAT 1646

Qy       1593 GAATGAAGATGGTGGTTTTGCAACATATGAGCTTAAACGATCTTATAATTGGTTGGAGAT 1652
              ||||||  ||||||| |||||||||||||| || | | | ||||||   ||| ||||| |
Db       1647 GAATGATAATGGTGGATTTGCAACATATGAACTCACAAGGTCTTATGCCTGGCTGGAGCT 1706

Qy       1653 AATCAATCCTGCTGAAACTTTTGGTGACATCGTTATTGATTATCCTTATGTGGAATGTAC 1712
              |||||||||||||| || ||||| ||||| || ||||||||||||||||||||||| ||
Db       1707 TATCAATCCTGCTGAGACCTTTGGGGACATTGTGATTGATTATCCTTATGTGGAATGCAC 1766

Qy       1713 ATCAGCAGCGATTCAAGCTTTGGCATCATTTAGGAAATTATATCCTGGGCATCGCCGAGA 1772
              |||||||| ||||| ||  || || ||||||  ||  | || |||||  || |||
Db       1767 TTCAGCAGCAATTCAGGCCCTGACAGCATTTAAAAAGCTCTACCCTGGACACCGCAAGAG 1826

Qy       1773 AGAAATACAACATTGTATCGATAAAGCCACTACCTTCATTGAAAAAATACAAGCTTCAGA 1832
              ||||||  |  |||||     ||||| ||| ||| |||||     || |||      ||
Db       1827 TGAAATAGACAACTGTATAAGCAAAGCTGCTAGCTTTATTGAGGGTATTCAAAAAAGCGA 1886

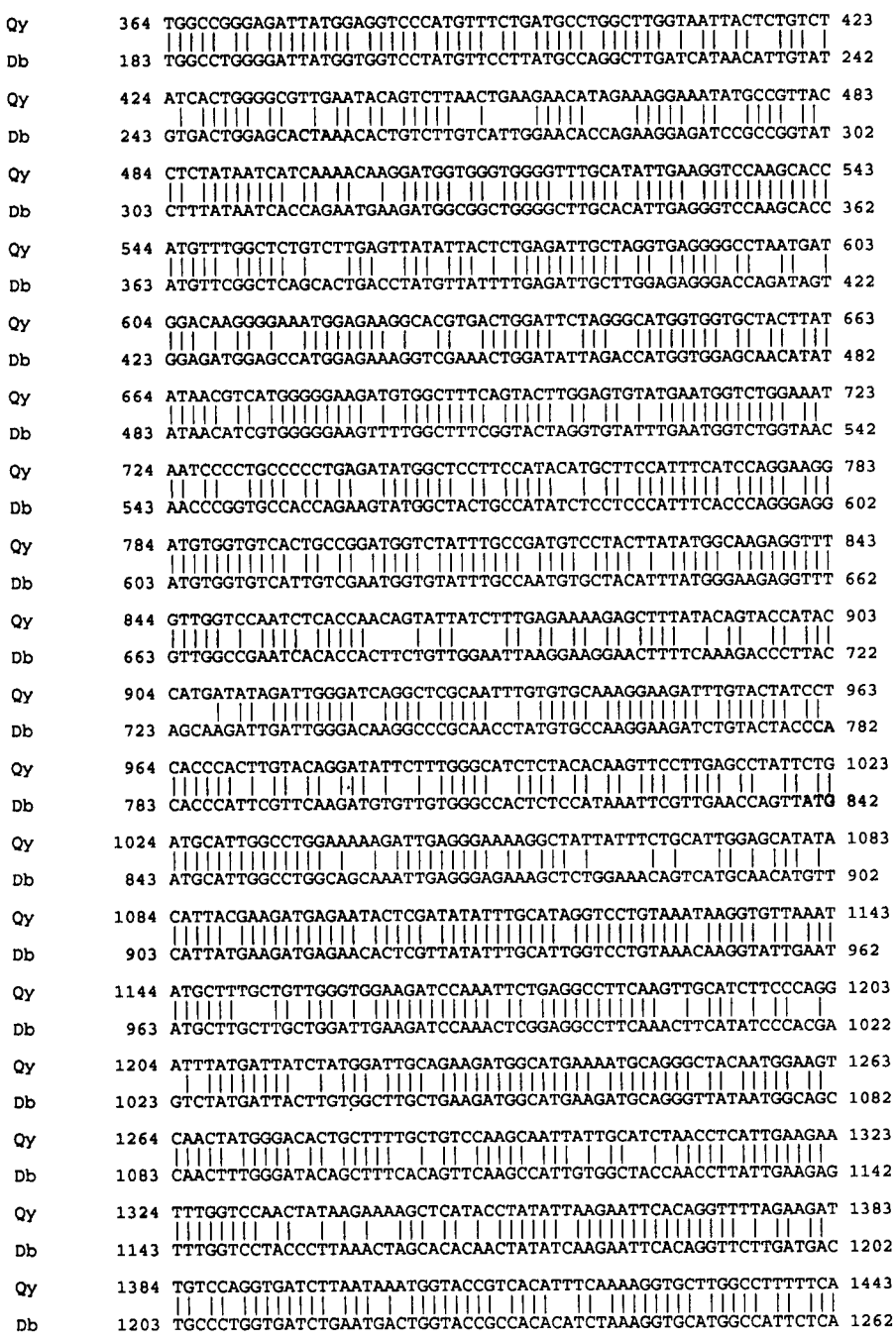
Qy       1833 TGGATCATGGTATGGTTCTTGGGGAGTTTGCTTCACTTACGGTGCTTGGTTTGGGGTAAA 1892
              ||| ||||||||||||||||||  ||||| || || || ||  | ||||||| |||||
Db       1887 TGGTTCATGGTATGGTTCTTGGGCTGTTTGTTTTACCTATGGCACATGGTTTGGTGTAAA 1946

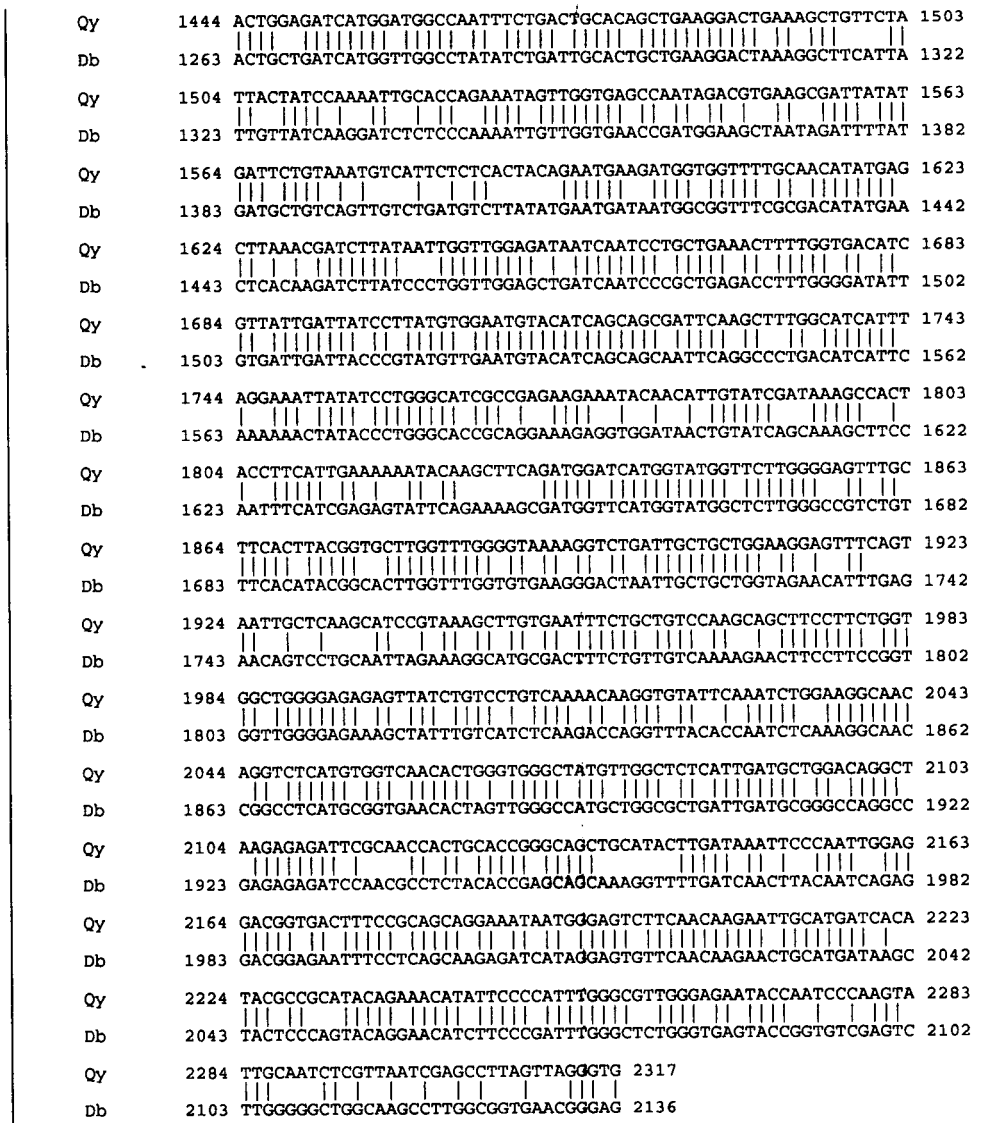
Qy       1893 AGGTCTGATTGCTGCTGGAAGGAGTTTCAGTAATTGCTCAAGCATCCGTAAAGCTTGTGA 1952
              ||  |  |||||||||| ||||  ||||  ||  |  |    ||| | || || |||||
Db       1947 GGGATTAGTTGCTGCTGGTAGGACATTCAAAAACAGTCCTGCAATCAGAAAGGCATGTGA 2006

Qy       1953 ATTTCTGCTGTCCAAGCAGCTTCCTTCTGGTGGCTGGGGAGAGAGTTATCTGTCCTGTCA 2012
              ||| || |||| ||  ||||| |||||||||| || ||| |||| |  ||
Db       2007 CTTTTTGTTGTCAAAAGAGCTTCCTTCTGGAGGCTGGGGAGAAAGCTATTTGTCATCCCA 2066

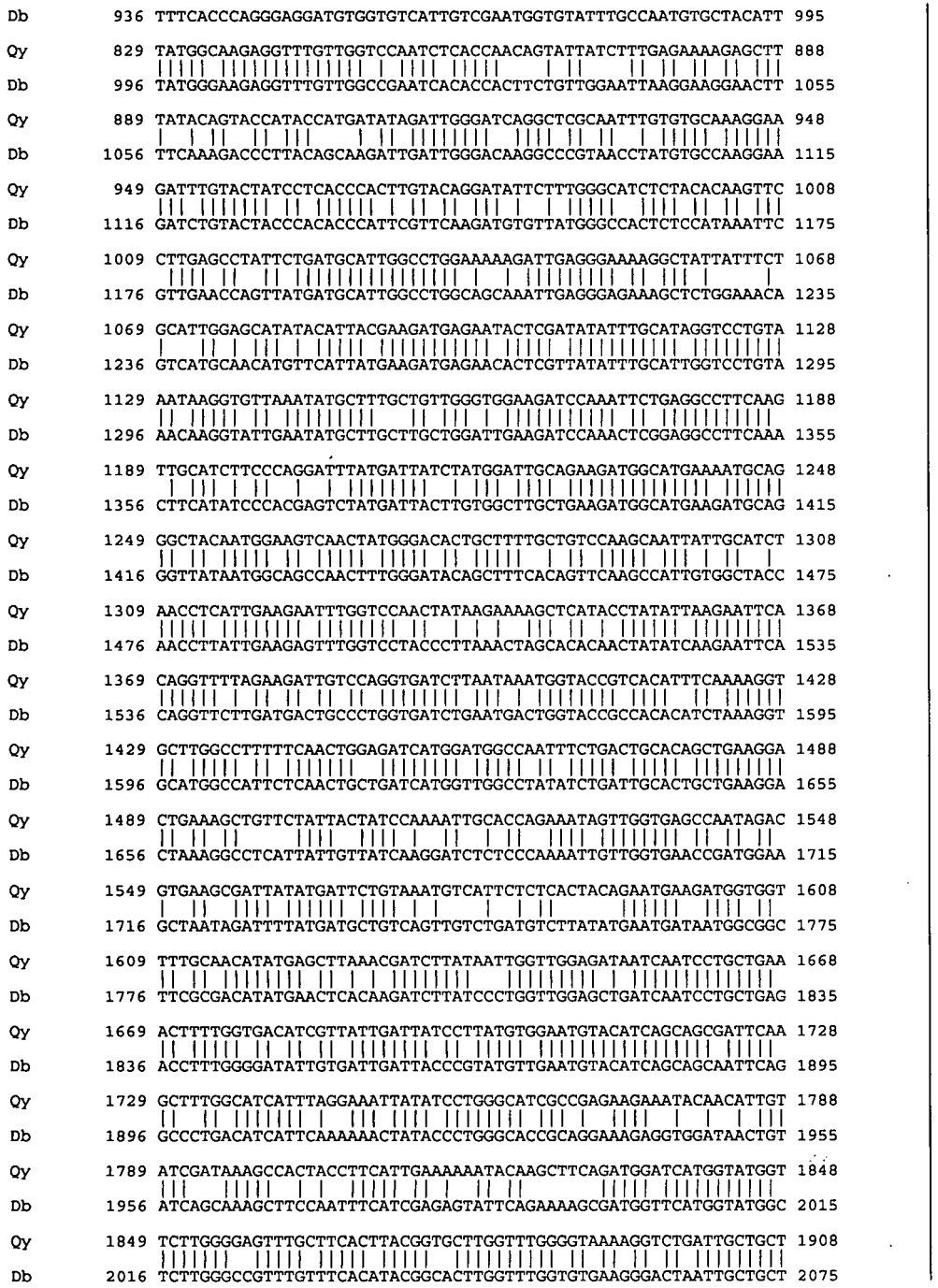
Qy       2013 AAACAAGGTGTATTCAAATCTGGAAGGCAACAGGTCTCATGTGGTCAACACTGGGTGGGC 2072
              | |  |||| ||| | ||||| ||  |  |||||| ||| |||||||| |||||
Db       2067 AGATCAGGTTTATACCAATCTCGAAGGGAAGCGACCTCATGCGGTGAACACTGGTTGGGC 2126

Qy       2073 TATGTTGGCTCTCATTGATGCTGGACAGGCTAAGAGAGATTCGCAACCACTGCACCGGGC 2132
              ||| | || || || ||||| || |||||| |||||||| |     ||  |||| || ||
Db       2127 CATGCTAGCCCTAATCGATGCAGGGCAGGCTGAGAGAGATCCAATTCCTTTGCATCGAGC 2186
```





RESULT 9
US-10-437-963-71159
; Sequence 71159, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad



RESULT 8
US-10-607-726-1
; Sequence 1, Application US/10607726
; Publication No. US20040121439A1
; GENERAL INFORMATION:
; APPLICANT: Falco, S. Carl
; APPLICANT: Pamodu, Omolayo O.
; APPLICANT: Hitz, William D.
; APPLICANT: Kinney, Anthony J.
; APPLICANT: Rafalski, Antoni
; APPLICANT: McGonigle, Brian
; APPLICANT: Lohman, Karin
; TITLE OF INVENTION: Enzymes Involved In Squalene Metabolizm
; FILE REFERENCE: BB1113 US NA
; CURRENT APPLICATION NUMBER: US/10/607,726
; CURRENT FILING DATE: 2003-06-27
; PRIOR APPLICATION NUMBER: US/09/419,679
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/105,405
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-10-23
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 1
; LENGTH: 2558
; TYPE: DNA
; ORGANISM: Zea mays
US-10-607-726-1

Query Match 46.2%; Score 1183.6; DB 20; Length 2558;
Best Local Similarity 72.2%; Pred. No. 7.9e-314;
Matches 1540; Conservative 0; Mismatches 594; Indels 0; Gaps 0;

```
Qy       1309 AACCTCATTGAAGAATTTGGTCCAACTATAAGAAAAGCTCATACCTATATTAAGAATTCA 1368
Db       1476 AACCTTATTGAAGAGTTTGGTCCTACCCTTAAACTAGCACACAACTATATCAAGAATTCA 1535

Qy       1369 CAGGTTTTAGAAGATTGTCCAGGTGATCTTAATAAATGGTACCGTCACATTTCAAAAGGT 1428
Db       1536 CAGGTTCTTGATGACTGCCCTGGTGATCTGAATGACTGGTACCGCCACACATCTAAAGGT 1595

Qy       1429 GCTTGGCCTTTTTCAACTGGAGATCATGGATGGCCAATTTCTGACTGCACAGCTGAAGGA 1488
Db       1596 GCATGGCCATTCTCAACTGCTGATCATGGTTGGCCTATATCTGATTGCACTGCTGAAGGA 1655

Qy       1489 CTGAAAGCTGTTCTATTACTATCCAAAATTGCACCAGAAATAGTTGGTGAGCCAATAGAC 1548
Db       1656 CTAAAGGCCTCATTATTGTTATCAAGGATCTCTCCCAAAATTGTTGGTGAACCGATGGAA 1715

Qy       1549 GTGAAGCGATTATATGATTCTGTAAATGTCATTCTCTCACTACAGAATGAAGATGGTGGT 1608
Db       1716 GCTAATAGATTTTATGATGCTGTCAGTTGTCTGATGTCTTATATGAATGATAATGGCGGC 1775

Qy       1609 TTTGCAACATATGAGCTTAAACGATCTTATAATTGGTTGGAGATAATCAATCCTGCTGAA 1668
Db       1776 TTCGCGACATATGAACTCACAAGATCTTATCCCTGGTTGGAGCTGATCAATCCTGCTGAG 1835

Qy       1669 ACTTTTGGTGACATCGTTATTGATTATCCTTATGTGGAATGTACATCAGCAGCGATTCAA 1728
Db       1836 ACCTTTGGGGATATTGTGATTGATTACCCGTATGTTGAATGTACATCAGCAGCAATTCAG 1895

Qy       1729 GCTTTGGCATCATTTAGGAAATTATATCCTGGGCATCGCCGAGAAGAAATACAACATTGT 1788
Db       1896 GCCCTGACATCATTCAAAAAACTATACCCTGGGCACCGCAGGAAAGAGGTGGATAACTGT 1955

Qy       1789 ATCGATAAAGCCACTACCTTCATTGAAAAAATACAAGCTTCAGATGGATCATGGTATGGT 1848
Db       1956 ATCAGCAAAGCTTCCAATTTCATCGAGAGTATTCAGAAAAGCGATGGTTCATGGTATGGC 2015

Qy       1849 TCTTGGGGAGTTTGCTTCACTTACGGTGCTTGGTTTGGGGTAAAAGGTCTGATTGCTGCT 1908
Db       2016 TCTTGGGCCGTTTGTTTCACATACGGCACTTGGTTTGGTGTGAAGGGACTAATTGCTGCT 2075

Qy       1909 GGAAGGAGTTTCAGTAATTGCTCAAGCATCCGTAAAGCTTGTGAATTTCTGCTGTCCAAG 1968
Db       2076 GGTAGAACATTTGAGAACAGTCCTGCAATTAGAAAGGCATGCGACTTTCTGTTGTCAAAA 2135

Qy       1969 CAGCTTCCTTCTGGTGGCTGGGGAGAGAGTTATCTGTCCTGTCAAAACAAGGTGTATTCA 2028
Db       2136 GAACTTCCTTCCGGTGGTTGGGGAGAAAGCTATTTGTCATCTCAAGACCAGGTTTACACC 2195

Qy       2029 AATCTGGAAGGCAACAGGTCTCATGTGGTCAACACTGGGTGGGCTATGTTGGCTCTCATT 2088
Db       2196 AATCTCAAAGGCAACCGGCCTCATGCAGTGAACACTAGTTGGGCCATGCTGGCGCTGATT 2255

Qy       2089 GATGCTGGACAGGCTAAGAGAGATTCGCAACCACTGCACCGGGCAGCTGCATACTTGATA 2148
Db       2256 GATGCAGGCCAGGCCGAGAGAGATCCAACGCCTCTACACCGAGCAGCAAAGGTTTTGATC 2315

Qy       2149 AATTCCCAATTGGAGGACGGTGACTTTCCGCAGCAGGAAATAATGGGAGTCTTCAACAAG 2208
Db       2316 AACTTGCAATCACATGATGGAGAATTTCCTCAGCAAGAGATCATAGGAGTGTTCAACAAG 2375

Qy       2209 AATTGCATGATCACATACGCCGCATACAGAAACATATTCCCCATTTGGGCGTTGGGAGAA 2268
Db       2376 AACTGCATGATAAGCTACTCCCAGTACAGGAACATCTTCCCGATTTGGGCTCTGGGTGAG 2435

Qy       2269 TACCAATCCCAAGTATTGCAATCTCGTTAATCGAGCCTTAGTTAGGGTG 2317
Db       2436 TACCGGTGCCGAGTCTTGGGGGCTGGCAAGTCTTGGCGGTGAACGGGAG 2484

RESULT 7
US-10-425-115-90182
; Sequence 90182, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(53222)B
;  CURRENT APPLICATION NUMBER: US/10/425,115
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 90182
;   LENGTH: 2786
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: MRT4577_182249C.1
US-10-425-115-90182

  Query Match             46.7%;  Score 1196.2;  DB 21;  Length 2786;
  Best Local Similarity   70.8%;  Pred. No. 0;
  Matches 1607;  Conservative    0;  Mismatches  653;  Indels    9;  Gaps    1;

Qy         49 GGAGGGAATCCATGGCTTCGGACATTGAACAATCACGTTGGAAGACAGGTGTGGGAGTTC 108
Db        225 GGCGGGGACCCCTGGCTACGCACGAAGAACGCCCACGTCGGCCGCCAGGTGTGGGAGTTC 284

Qy        109 GATCCTAAGCTTGGATCGCCGCAAGATCTCCTCGAGATTGAGAAAGCTCGCCAGAATTTT 168
Db        285 GACGCCG---------CCGCCGACCCCGACCCCGCCGTCGACGCCGCGCGCCGCGCCTTC 335

Qy        169 CACGATAACCGCTTTACCCACAAACACAGCGCTGATCTACTTATGCGGATGCAGTTCGCA 228
Db        336 GTCGCCACCAGGGGCCACCTCAAGCACAGCGCCGACCTCCTCATGCGCATCCAGTTCGCC 395

Qy        229 AGAGAGAACCCAACACGTGAAGTCTTGCCCAAAGTCGGAGTTAAGGATATTGAGGATGTG 288
Db        396 AAAGAAAACTCGATTGAGCTTCACCTTCCAGGCATCAAGCTCGGTGAGCATGAAGATGTG 455

Qy        289 ACCCAAGAGATTGTGACAAAAACATTAAGAAGGGCCGTAAGTTTCCATTCAACTCTCCAG 348
Db        456 ACCGAGGAAGCTGTGTTGACTACATTGAAAAGGGCAATCAGCCGTTTCTCTACTCTCCAG 515

Qy        349 TGCCATGACGGACACTGGCCGGGAGATTATGGAGGTCCCATGTTTCTGATGCCTGGCTTG 408
Db        516 GCACATGATGGACACTGGCCTGGGGATTATGGTGGTCCTATGTTCCTTATGCCAGGCTTG 575

Qy        409 GTAATTACTCTGTCTATCACTGGGGCGTTGAATACAGTCTTAACTGAAGAACATAGAAAG 468
Db        576 ATCATAACATTGTATGTGACTGGAGCACTAAACACTGTCTTGTCATTGGAACACCAGAAG 635

Qy        469 GAAATATGCCGTTACCTCTATAATCATCAAAACAAGGATGGTGGGTGGGGTTTGCATATT 528
Db        636 GAGATCCGCCGGTATCTTTATAATCACCAGAATGAAGATGGCGGCTGGGGCTTGCACATT 695

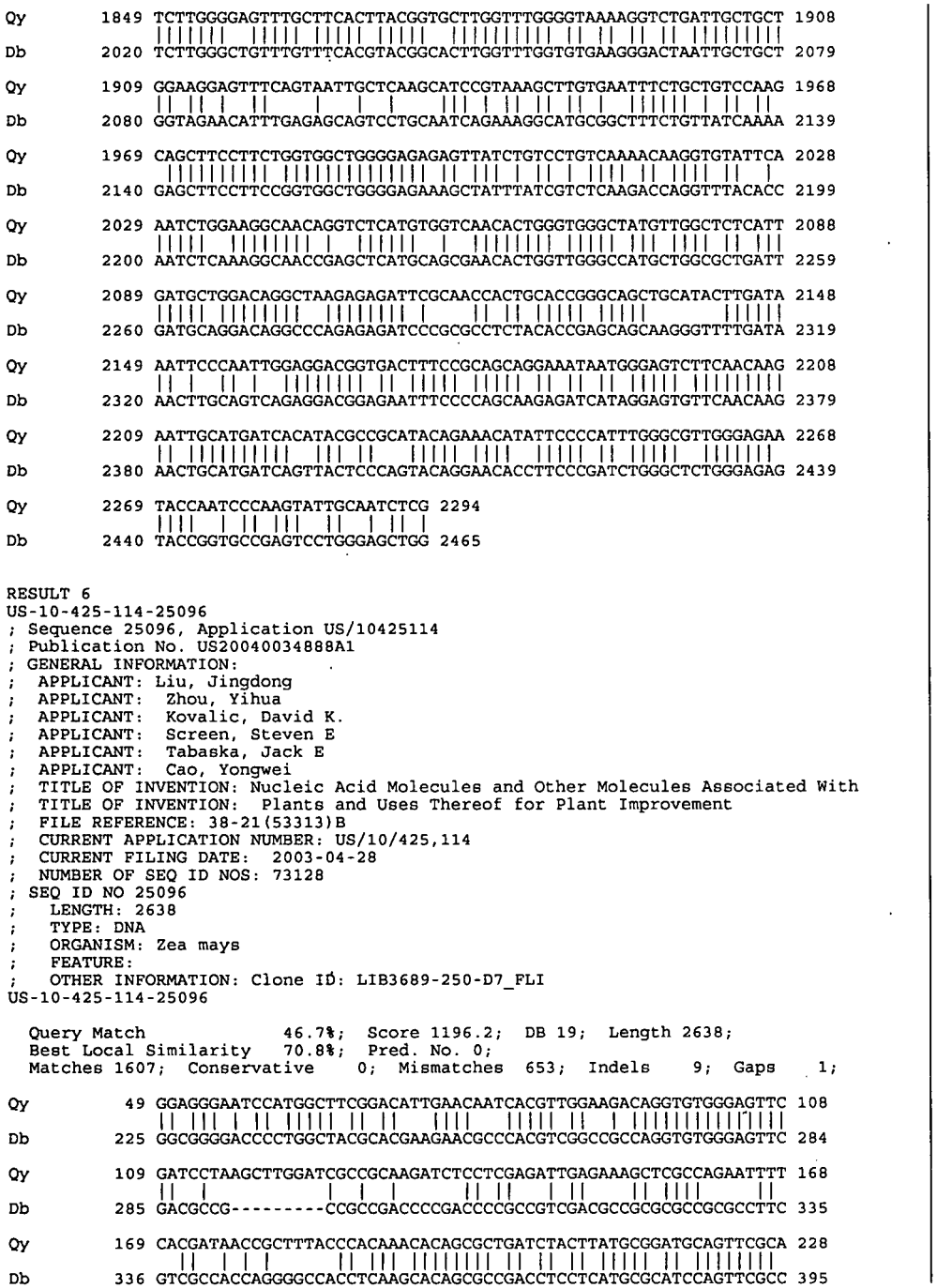
Qy        529 GAAGGTCCAAGCACCATGTTTGGCTCTGTCTTGAGTTATATTACTCTGAGATTGCTAGGT 588
Db        696 GAGGGTCCAAGCACCATGTTCGGCTCAGCACTGACCTATGTTATTTTGAGATTGCTTGGA 755

Qy        589 GAGGGGCCTAATGATGGACAAGGGGAAATGGAGAAGGCACGTGACTGGATTCTAGGGCAT 648
Db        756 GAGGGACCAGATAGTGGAGATGGAGCCATGGAGAAAGGTCGAAACTGGATATTAGACCAT 815

Qy        649 GGTGGTGCTACTTATATAACGTCATGGGGGAAGATGTGGCTTTCAGTACTTGGAGTGTAT 708
Db        816 GGTGGAGCAACATATATAACATCGTGGGGGAAGTTTTGGCTTTCGGTACTAGGTGTATTT 875

Qy        709 GAATGGTCTGGAAATAATCCCCTGCCCCCTGAGATATGGCTCCTTCCATACATGCTTCCA 768
Db        876 GAATGGTCTGGTAACAACCCGGTGCCACCAGAAGTATGGCTACTGCCATATCTCCTCCCA 935

Qy        769 TTTCATCCAGGAAGGATGTGGTGTCACTGCCGGATGGTCTATTTGCCGATGTCCTACTTA 828
```



```
Qy       1849 TCTTGGGGAGTTTGCTTCACTTACGGTGCTTGGTTTGGGGTAAAAGGTCTGATTGCTGCT 1908
Db       2020 TCTTGGGCTGTTTGTTTCACGTACGGCACTTGGTTTGGTGTGAAGGGACTAATTGCTGCT 2079

Qy       1909 GGAAGGAGTTTCAGTAATTGCTCAAGCATCCGTAAAGCTTGTGAATTTCTGCTGTCCAAG 1968
Db       2080 GGTAGAACATTTGAGAGCAGTCCTGCAATCAGAAAGGCATGCGGCTTTCTGTTATCAAAA 2139

Qy       1969 CAGCTTCCTTCTGGTGGCTGGGGAGAGAGTTATCTGTCCTGTCAAAACAAGGTGTATTCA 2028
Db       2140 GAGCTTCCTTCCGGTGGCTGGGGAGAAAGCTATTTATCGTCTCAAGACCAGGTTTACACC 2199

Qy       2029 AATCTGGAAGGCAACAGGTCTCATGTGGTCAACACTGGGTGGGCTATGTTGGCTCTCATT 2088
Db       2200 AATCTCAAAGGCAACCGAGCTCATGCAGCGAACACTGGTTGGGCCATGCTGGCGCTGATT 2259

Qy       2089 GATGCTGGACAGGCTAAGAGAGATTCGCAACCACTGCACCGGGCAGCTGCATACTTGATA 2148
Db       2260 GATGCAGGACAGGCCCAGAGAGATCCCGCGCCTCTACACCGAGCAGCAAGGGTTTTGATA 2319

Qy       2149 AATTCCCAATTGGAGGACGGTGACTTTCCGCAGCAGGAAATAATGGGAGTCTTCAACAAG 2208
Db       2320 AACTTGCAGTCAGAGGACGGAGAATTTCCCCAGCAAGAGATCATAGGAGTGTTCAACAAG 2379

Qy       2209 AATTGCATGATCACATACGCCGCATACAGAAACATATTCCCCATTTGGGCGTTGGGAGAA 2268
Db       2380 AACTGCATGATCAGTTACTCCCAGTACAGGAACACCTTCCCGATCTGGGCTCTGGGAGAG 2439

Qy       2269 TACCAATCCCAAGTATTGCAATCTCG 2294
Db       2440 TACCGGTGCCGAGTCCTGGGAGCTGG 2465


RESULT 6
US-10-425-114-25096
; Sequence 25096, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 25096
;   LENGTH: 2638
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: LIB3689-250-D7_FLI
US-10-425-114-25096

  Query Match             46.7%;  Score 1196.2;  DB 19;  Length 2638;
  Best Local Similarity   70.8%;  Pred. No. 0;
  Matches 1607;  Conservative    0;  Mismatches  653;  Indels    9;  Gaps    1;

Qy         49 GGAGGGAATCCATGGCTTCGGACATTGAACAATCACGTTGGAAGACAGGTGTGGGAGTTC 108
Db        225 GGCGGGGACCCCTGGCTACGCACGAAGAACGCCCACGTCGGCCGCCAGGTGTGGGAGTTC 284

Qy        109 GATCCTAAGCTTGGATCGCCGCAAGATCTCCTCGAGATTGAGAAAGCTCGCCAGAATTTT 168
Db        285 GACGCCG---------CCGCCGACCCCGACCCCGCCGTCGACGCCGCGCGCCGCGCCTTC 335

Qy        169 CACGATAACCGCTTTACCCACAAACACAGCGCTGATCTACTTATGCGGATGCAGTTCGCA 228
Db        336 GTCGCCACCAGGGGCCACCTCAAGCACAGCGCCGACCTCCTCATGCGCATCCAGTTCGCC 395

Qy        229 AGAGAGAACCCAACACGTGAAGTCTTGCCCAAAGTCGGAGTTAAGGATATTGAGGATGTG 288
Db        396 AAAGAAAACTCGATTGAGCTTCACCTTCCAGGCATCAAGCTCGGTGAGCATGAAGATGTG 455

Qy        289 ACCCAAGAGATTGTGACAAAAACATTAAGAAGGGCCGTAAGTTTCCATTCAACTCTCCAG 348
Db        456 ACCGAGGAAGCTGTGTTGACTACATTGAAAAGGGCAATCAGCCGTTTCTCTACTCTCCAG 515

Qy        349 TGCCATGACGGACACTGGCCGGGAGATTATGGAGGTCCCATGTTTCTGATGCCTGGCTTG 408
Db        516 GCACATGATGGACACTGGCCTGGGGATTATGGTGGTCCTATGTTCCTTATGCCAGGCTTG 575

Qy        409 GTAATTACTCTGTCTATCACTGGGGCGTTGAATACAGTCTTAACTGAAGAACATAGAAAG 468
Db        576 ATCATAACATTGTATGTGACTGGAGCACTAAACACTGTCTTGTCATTGGAACACCAGAAG 635

Qy        469 GAAATATGCCGTTACCTCTATAATCATCAAAACAAGGATGGTGGGTGGGGTTTGCATATT 528
Db        636 GAGATCCGCCGGTATCTTTATAATCACCAGAATGAAGATGGCGGCTGGGGCTTGCACATT 695

Qy        529 GAAGGTCCAAGCACCATGTTTGGCTCTGTCTTGAGTTATATTACTCTGAGATTGCTAGGT 588
Db        696 GAGGGTCCAAGCACCATGTTCGGCTCAGCACTGACCTATGTTATTTTGAGATTGCTTGGA 755

Qy        589 GAGGGGCCTAATGATGGACAAGGGGAAATGGAGAAGGCACGTGACTGGATTCTAGGGCAT 648
Db        756 GAGGGACCAGATAGTGGAGATGGAGCCATGGAGAAAGGTCGGAACTGGATATTAGACCAT 815

Qy        649 GGTGGTGCTACTTATATAACGTCATGGGGGAAGATGTGGCTTTCAGTACTTGGAGTGTAT 708
Db        816 GGTGGAGCAACATATATAACATCGTGGGGGAAGTTTTGGCTTTCGGTACTAGGTGTATTT 875

Qy        709 GAATGGTCTGGAAATAATCCCCTGCCCCCTGAGATATGGCTCCTTCCATACATGCTTCCA 768
Db        876 GAATGGTCTGGTAACAACCCGGTGCCACCAGAAGTATGGCTACTGCCATATCTCCTCCCA 935

Qy        769 TTTCATCCAGGAAGGATGTGGTGTCACTGCCGGATGGTCTATTTGCCGATGTCCTACTTA 828
Db        936 TTTCACCCAGGGAGGATGTGGTGTCATTGTCGAATGGTGTATTTGCCAATGTGCTACATT 995

Qy        829 TATGGCAAGAGGTTTGTTGGTCCAATCTCACCAACAGTATTATCTTTGAGAAAAGAGCTT 888
Db        996 TATGGGAAGAGGTTTGTTGGCCGAATCACACCACTTCTGTTGGAATTAAGGAAGGAACTT 1055

Qy        889 TATACAGTACCATACCATGATATAGATTGGGATCAGGCTCGCAATTTGTGTGCAAAGGAA 948
Db       1056 TTCAAAGACCCTTACAGCAAGATTGATTGGGACAAGGCCCGTAACCTATGTGCCAAGGAA 1115

Qy        949 GATTTGTACTATCCTCACCCACTTGTACAGGATATTCTTTGGGCATCTCTACACAAGTTC 1008
Db       1116 GATCTGTACTACCCACACCCATTCGTTCAAGATGTGTTATGGGCCACTCTCCATAAATTC 1175

Qy       1009 CTTGAGCCTATTCTGATGCATTGGCCTGGAAAAAGATTGAGGGAAAAGGCTATTATTTCT 1068
Db       1176 GTTGAACCAGTTATGATGCATTGGCCTGGCAGCAAATTGAGGGAGAAAGCTCTGGAAACA 1235

Qy       1069 GCATTGGAGCATATACATTACGAAGATGAGAATACTCGATATATTTGCATAGGTCCTGTA 1128
Db       1236 GTCATGCAACATGTTCATTATGAAGATGAGAACACTCGTTATATTTGCATTGGTCCTGTA 1295

Qy       1129 AATAAGGTGTTAAATATGCTTTGCTGTTGGGTGGAAGATCCAAATTCTGAGGCCTTCAAG 1188
Db       1296 AACAAGGTATTGAATATGCTTGCTTGCTGGATTGAAGATCCAAACTCGGAGGCCTTCAAA 1355

Qy       1189 TTGCATCTTCCCAGGATTTATGATTATCTATGGATTGCAGAAGATGGCATGAAAATGCAG 1248
Db       1356 CTTCATATCCCACGAGTCTATGATTACTTGTGGCTTGCTGAAGATGGCATGAAGATGCAG 1415

Qy       1249 GGCTACAATGGAAGTCAACTATGGGACACTGCTTTTGCTGTCCAAGCAATTATTGCATCT 1308
Db       1416 GGTTATAATGGCAGCCAACTTTGGGATACAGCTTTCACAGTTCAAGCCATTGTGGCTACC 1475
```

US-10-425-115-91064
; Sequence 91064, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 91064
; LENGTH: 2776
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_183051C.1
US-10-425-115-91064

Query Match 46.8%; Score 1197.2; DB 21; Length 2776;
Best Local Similarity 71.2%; Pred. No. 0;
Matches 1599; Conservative 0; Mismatches 638; Indels 9; Gaps 1;

```
Qy        49 GGAGGGAATCCATGGCTTCGGACATTGAACAATCACGTTGGAAGACAGGTGTGGGAGTTC 108
Db       229 GGCGGGGACCCCTGGCTCCGCACGAAGAACGCCCACGTCGGCCGCCAGGTCTGGGAGTTC 288

Qy       109 GATCCTAAGCTTGGATCGCCGCAAGATCTCCTCGAGATTGAGAAAGCTCGCCAGAATTTT 168
Db       289 GACGCC---------TCCGCCGACCCCGACCCCGCCGTCGACGCCGCGCGCCGCGCCTTC 339

Qy       169 CACGATAACCGCTTTACCCACAAACACAGCGCTGATCTACTTATGCGGATGCAGTTCGCA 228
Db       340 GCCGGCAGCAGGGGCCACCTCAAGCACAGCGCCGACCTCCTCATGCGCATCCAGTTCGCC 399

Qy       229 AGAGAGAACCCAACACGTGAAGTCTTGCCCAAAGTCGGAGTTAAGGATATTGAGGATGTG 288
Db       400 AAAGAAAATCCTCTTGAGCTTGACCTTCCAGCGATCAAGCTCGGTGAGCACGAAGACGTA 459

Qy       289 ACCCAAGAGATTGTGACAAAAACATTAAGAAGGGCCGTAAGTTTCCATTCAACTCTCCAG 348
Db       460 ACCGAGGAAGCCGTGTCTACTACTTTGAGAAGGGCAATCAGCCGTCTTTCTACTCTCCAG 519

Qy       349 TGCCATGACGGACACTGGCCGGGAGATTATGGAGGTCCCATGTTTCTGATGCCTGGCTTG 408
Db       520 GCACATGATGGACACTGGCCTGGGGATTATGGTGGCCCTATGTTCCTTATGCCAGGCTTG 579

Qy       409 GTAATTACTCTGTCTATCACTGGGGCGTTGAATACAGTCTTAACTGAAGAACATAGAAAG 468
Db       580 ATCATAACATTGTATGTGACTGGAGCACTAAACACTGTCTTGTCATCAGAACACCAGAAG 639

Qy       469 GAAATATGCCGTTACCTCTATAATCATCAAAACAAGGATGGTGGGTGGGGTTTGCATATT 528
Db       640 GAGATCCGTCGGTATCTTTATAATCATCAGAATGAAGATGGTGGCTGGGGCTTGCACATT 699

Qy       529 GAAGGTCCAAGCACCATGTTTGGCTCTGTCTTGAGTTATATTACTCTGAGATTGCTAGGT 588
Db       700 GAGGGTCCAAGCACCATGTTTGGCTCAGCACTGACCTACGTTACTTTGAGATTGCTTGGA 759

Qy       589 GAGGGGCCTAATGATGGACAAGGGGAAATGGAGAAGGCACGTGACTGGATTCTAGGGCAT 648
Db       760 GAGGGACCAGATAGTGGAGATGGAGCAATGGAGAAAGGCCGAAACTGGATCTTAGACCAT 819

Qy       649 GGTGGTGCTACTTATATAACGTCATGGGGGAAGATGTGGCTTTCAGTACTTGGAGTGTAT 708
Db       820 GGCGGAGCAACATATATAACATCTTGGGGGAAGTTTTGGCTCTCGGTACTTGGTGTATTT 879

Qy       709 GAATGGTCTGGAAATAATCCCCTGCCCCCTGAGATATGGCTCCTTCCATACATGCTTCCA 768
Db       880 GAATGGTCTGGTAACAACCCGGTACCACCAGAAGTATGGCTACTGCCATATCTCCTGCCA 939

Qy       769 TTTCATCCAGGAAGGATGTGGTGTCACTGCCGGATGGTCTATTTGCCGATGTCCTACTTA 828
Db       940 TTTCACCCAGGGAGGATGTGGTGTCATTGTCGAATGGTGTATTTGCCAATGTGCTACATC 999

Qy       829 TATGGCAAGAGGTTTGTTGGTCCAATCTCACCAACAGTATTATCTTTGAGAAAAGAGCTT 888
Db      1000 TATGGGAAGAGGTTTGTTGGTCGAATCACACCACTCGTGTTGGAATTAAGAAAGGAACTT 1059

Qy       889 TATACAGTACCATACCATGATATAGATTGGGATCAGGCTCGCAATTTGTGTGCAAAGGAA 948
Db      1060 TTCAAAGACCCTTACAGCAAGATTGATTGGGACAAGGCCCGCAATCTATGTGCTAAGGAA 1119

Qy       949 GATTTGTACTATCCTCACCCACTTGTACAGGATATTCTTTGGGCATCTCTACACAAGTTC 1008
Db      1120 GATCTGTACTATCCACATCCATTTGTTCAAGATGTCTTATGGGCCACTCTCCATAAATTT 1179

Qy      1009 CTTGAGCCTATTCTGATGCATTGGCCTGGAAAAAGATTGAGGGAAAAGGCTATTATTTCT 1068
Db      1180 GTTGAACCAGTTATGATGAGTTGGCCTGGCAGTAAATTGAGGGAGAAAGCTCTGGAGACA 1239

Qy      1069 GCATTGGAGCATATACATTACGAAGATGAGAATACTCGATATATTTGCATAGGTCCTGTA 1128
Db      1240 GCCATGCAACATGTTCACTATGAGGATGAGAACACTCGGTACATTTGCATTGGTCCTGTA 1299

Qy      1129 AATAAGGTGTTAAATATGCTTTGCTGTTGGGTGGAAGATCCAAATTCTGAGGCCTTCAAG 1188
Db      1300 AACAAGGTATTAAATATGCTTGCTTGCTGGATTGAAGATCCAAACTCAGAGGCCTTCAAA 1359

Qy      1189 TTGCATCTTCCCAGGATTTATGATTATCTATGGATTGCAGAAGATGGCATGAAAATGCAG 1248
Db      1360 CTCCACATCCCACGAGTCTATGACTACTTGTGGCTTGCTGAAGATGGCATGAAGATGCAG 1419

Qy      1249 GGCTACAATGGAAGTCAACTATGGGACACTGCTTTTGCTGTCCAAGCAATTATTGCATCT 1308
Db      1420 GGCTATAATGGTAGCCAACTTTGGGATACAGCTTTCATAGTTCAAGCCATTGTGGCTACC 1479

Qy      1309 AACCTCATTGAAGAATTTGGTCCAACTATAAGAAAAGCTCATACCTATATTAAGAATTCA 1368
Db      1480 AACCTTACCGAAGAGTTTGGTCCCACCCTTAAACTAGCACACAACTATATCAAGAAATCA 1539

Qy      1369 CAGGTTTTAGAAGATTGTCCAGGTGATCTTAATAAATGGTACCGTCACATTTCAAAAGGT 1428
Db      1540 CAGGTTCTTGACGACTGCCCTGGTGATCTGAATGACTGGTACCGCCACACATCTAAAGGC 1599

Qy      1429 GCTTGGCCTTTTTCAACTGGAGATCATGGATGGCCAATTTCTGACTGCACAGCTGAAGGA 1488
Db      1600 GCATGGCCATTCTCAACCGCGGATCATGGTTGGCCTATATCTGATTGCACTGCTGAAGGA 1659

Qy      1489 CTGAAAGCTGTTCTATTACTATCCAAAATTGCACCAGAAATAGTTGGTGAGCCAATAGAC 1548
Db      1660 CTAAAGGCTTCATTATTGTTATCAAAGATTTCTCCTGAAATTGTTGGTGGACCGATAGAA 1719

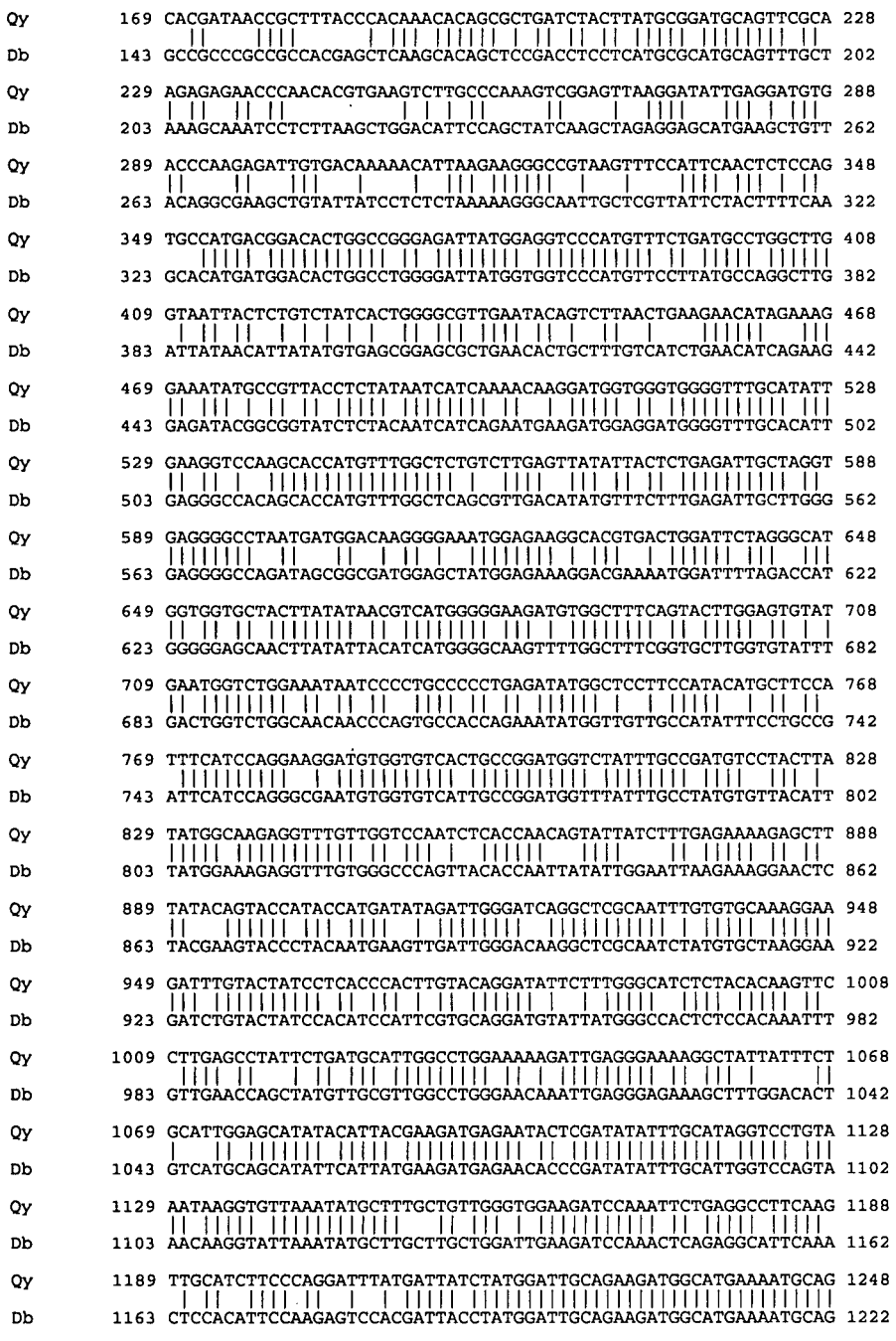
Qy      1549 GTGAAGCGATTATATGATTCTGTAAATGTCATTCTCTCACTACAGAATGAAGATGGTGGT 1608
Db      1720 GCTAATAGATTTTATGATGCTGTCAGTTGTCTGATGTCTTATATGAATGATAATGGTGGT 1779

Qy      1609 TTTGCAACATATGAGCTTAAACGATCTTATAATTGGTTGGAGATAATCAATCCTGCTGAA 1668
Db      1780 TTTGCGACTTATGAACTCACAAGATCCTATGCCTGGCTGGAGCTGATCAATCCTGCTGAG 1839

Qy      1669 ACTTTTGGTGACATCGTTATTGATTATCCTTATGTGGAATGTACATCAGCAGCGATTCAA 1728
Db      1840 ACCTTTGGGGATATTGTGATTGATTACCCGTATGTTGAATGTACATCAGCAGCAATCCAG 1899

Qy      1729 GCTTTGGCATCATTTAGGAAATTATATCCTGGGCATCGCCGAGAAGAAATACAACATTGT 1788
Db      1900 GCCCTGACATCATTCAGAAAACTATACCCTGGGCACCGCAGGAAAGAGGTGGACAACTGT 1959

Qy      1789 ATCGATAAAGCCACTACCTTCATTGAAAAAATACAAGCTTCAGATGGATCATGGTATGGT 1848
Db      1960 ATCAGCAAAGCTGCTAATTTCATTGAGAGCATTCAGAAAAGTGACGGTTCATGGTATGGC 2019
```



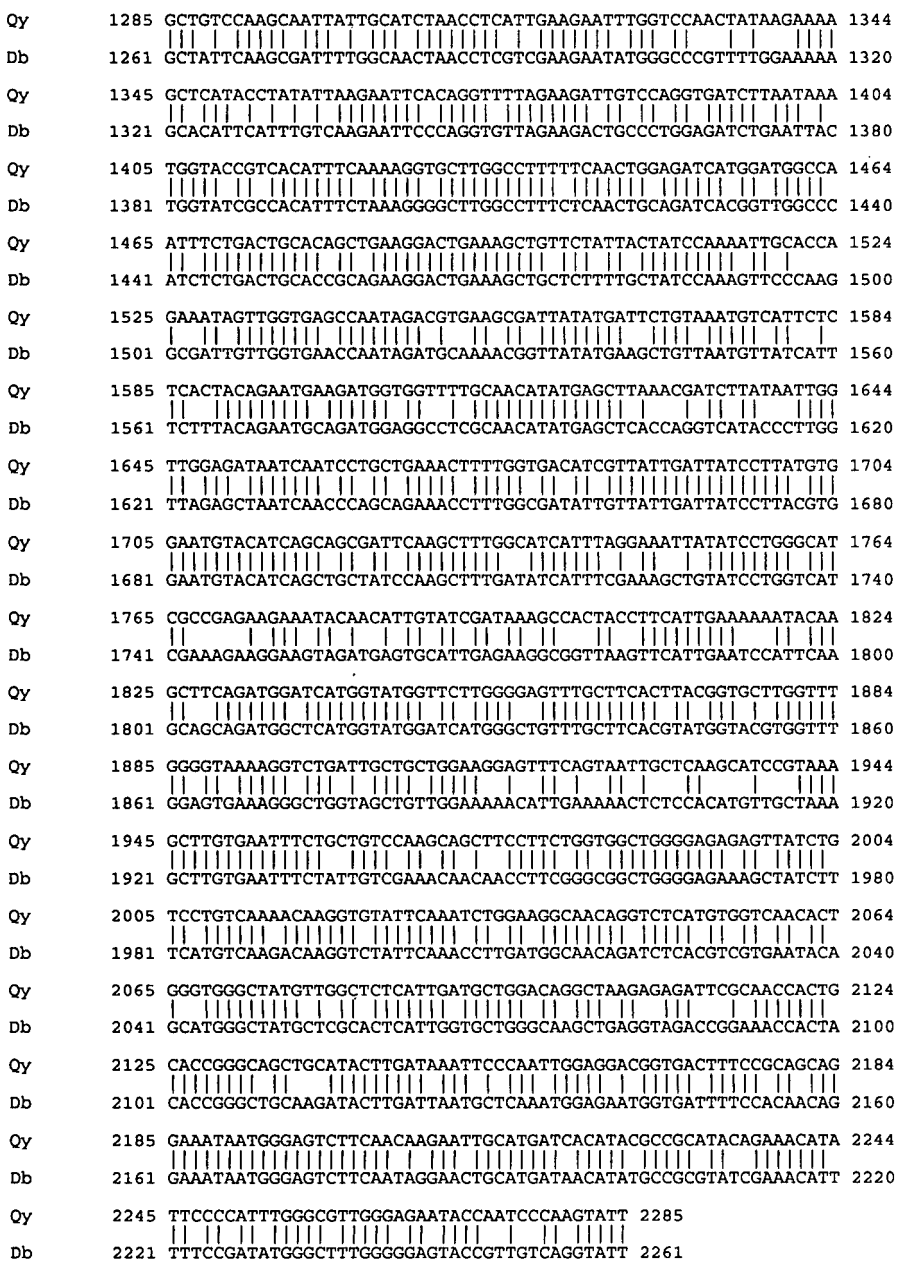
RESULT 5

```
Db        661 TGGCTATCGGTACTTGGAGCTTTTGAATGGTCCGGAAATAACCCACTGCCACCTGAGATA 720
Qy        745 TGGCTCCTTCCATACATGCTTCCATTTCATCCAGGAAGGATGTGGTGTCACTGCCGGATG 804
Db        721 TGGCTTCTCCCATATTTCCTGCCAATACATCCAGGAAGGATGTGGTGCCATTGTCGAATG 780
Qy        805 GTCTATTTGCCGATGTCCTACTTATATGGCAAGAGGTTTGTTGGTCCAATCTCACCAACA 864
Db        781 GTGTACTTGCCGATGTCGTATTTGTATGGAAAAAGGTTTGTGGGTCCCATAACGTCCACT 840
Qy        865 GTATTATCTTTGAGAAAAGAGCTTTATACAGTACCATACCATGATATAGATTGGGATCAG 924
Db        841 GTTTTATCACTGAGAAAGGAGCTTTTCACAGTACCATATCATGAAGTCAACTGGAATGAA 900
Qy        925 GCTCGCAATTTGTGTGCAAAGGAAGATTTGTACTATCCTCACCCACTTGTACAGGATATT 984
Db        901 GCACGCAACCTTTGCGCAAAGGAGGATTTATACTACCCACATCCACTTGTGCAAGATATT 960
Qy        985 CTTTGGGCATCTCTACACAAGTTCCTTGAGCCTATTCTGATGCATTGGCCTGGAAAAAGA 1044
Db        961 CTTTGGGCATCACTTCATAAGATTGTTGAGCCTGTTCTGATGCGATGGCCTGGTGCAAAT 1020
Qy       1045 TTGAGGGAAAAGGCTATTATTTCTGCATTGGAGCATATACATTACGAAGATGAGAATACT 1104
Db       1021 TTGAGAGAAAAGGCTATAAGAACCGCAATAGAACATATTCATTATGAAGATGAGAATACT 1080
Qy       1105 CGATATATTTGCATAGGTCCTGTAAATAAGGTGTTAAATATGCTTTGCTGTTGGGTGGAA 1164
Db       1081 AGGTACATCTGCATAGGTCCCGTGAACAAGGTATTAAATATGCTTTGCTGTTGGGTAGAA 1140
Qy       1165 GATCCAAATTCTGAGGCCTTCAAGTTGCATCTTCCCAGGATTTATGATTATCTATGGATT 1224
Db       1141 GACCCAAACTCAGAGGCTTTCAAGTTGCACCTACCAAGAATCCATGACTTTCTCTGGTTA 1200
Qy       1225 GCAGAAGATGGCATGAAAATGCAGGGCTACAATGGAAGTCAACTATGGGACACTGCTTTT 1284
Db       1201 GCTGAAGATGGAATGAAGATGCAGGGTTATAACGGAAGCCAGCTATGGGATACAGGTTTT 1260
Qy       1285 GCTGTCCAAGCAATTATTGCATCTAACCTCATTGAAGAATTTGGTCCAACTATAAGAAAA 1344
Db       1261 GCTATTCAAGCGATTTTGGCAACTAACCTCGTCGAAGAATATGGGCCCGTTTTGGAAAAA 1320
Qy       1345 GCTCATACCTATATTAAGAATTCACAGGTTTTAGAAGATTGTCCAGGTGATCTTAATAAA 1404
Db       1321 GCACATTCATTTGTCAAGAATTCCCAGGTGTTAGAAGACTGCCCTGGAGATCTGAATTAC 1380
Qy       1405 TGGTACCGTCACATTTCAAAAGGTGCTTGGCCTTTTTCAACTGGAGATCATGGATGGCCA 1464
Db       1381 TGGTATCGCCACATTTCTAAAGGGGCTTGGCCTTTCTCAACTGCAGATCACGGTTGGCCC 1440
Qy       1465 ATTTCTGACTGCACAGCTGAAGGACTGAAAGCTGTTCTATTACTATCCAAAATTGCACCA 1524
Db       1441 ATCTCTGACTGCACCGCAGAAGGACTGAAAGCTGCTCTTTTGCTATCCAAAGTTCCCAAG 1500
Qy       1525 GAAATAGTTGGTGAGCCAATAGACGTGAAGCGATTATATGATTCTGTAAATGTCATTCTC 1584
Db       1501 GCGATTGTTGGTGAACCAATAGATGCAAAACGGTTATATGAAGCTGTTAATGTTATCATT 1560
Qy       1585 TCACTACAGAATGAAGATGGTGGTTTTGCAACATATGAGCTTAAACGATCTTATAATTGG 1644
Db       1561 TCTTTACAGAATGCAGATGGAGGCCTCGCAACATATGAGCTCACCAGGTCATACCCTTGG 1620
Qy       1645 TTGGAGATAATCAATCCTGCTGAAACTTTTGGTGACATCGTTATTGATTATCCTTATGTG 1704
Db       1621 TTAGAGCTAATCAACCCAGCAGAAACCTTTGGCGATATTGTTATTGATTATCCTTACGTG 1680
Qy       1705 GAATGTACATCAGCAGCGATTCAAGCTTTGGCATCATTTAGGAAATTATATCCTGGGCAT 1764
Db       1681 GAATGTACATCAGCTGCTATCCAAGCTTTGATATCATTTCGAAAGCTGTATCCTGGTCAT 1740
Qy       1765 CGCCGAGAAGAAATACAACATTGTATCGATAAAGCCACTACCTTCATTGAAAAAATACAA 1824
Db       1741 CGAAAGAAGGAAGTAGATGAGTGCATTGAGAAGGCGGTTAAGTTCATTGAATCCATTCAA 1800
Qy       1825 GCTTCAGATGGATCATGGTATGGTTCTTGGGGAGTTTGCTTCACTTACGGTGCTTGGTTT 1884
Db       1801 GCAGCAGATGGCTCATGGTATGGATCATGGGCTGTTTGCTTCACGTATGGTACGTGGTTT 1860
Qy       1885 GGGGTAAAAGGTCTGATTGCTGCTGGAAGGAGTTTCAGTAATTGCTCAAGCATCCGTAAA 1944
Db       1861 GGAGTGAAAGGGCTGGTAGCTGTTGGAAAAACATTGAAAAACTCTCCACATGTTGCTAAA 1920
Qy       1945 GCTTGTGAATTTCTGCTGTCCAAGCAGCTTCCTTCTGGTGGCTGGGGAGAGAGTTATCTG 2004
Db       1921 GCTTGTGAATTTCTATTGTCGAAACAACAACCTTCGGGCGGCTGGGGAGAAAGCTATCTT 1980
Qy       2005 TCCTGTCAAAACAAGGTGTATTCAAATCTGGAAGGCAACAGGTCTCATGTGGTCAACACT 2064
Db       1981 TCATGTCAAGACAAGGTCTATTCAAACCTTGATGGCAACAGATCTCACGTCGTGAATACA 2040
Qy       2065 GGGTGGGCTATGTTGGCTCTCATTGATGCTGGACAGGCTAAGAGAGATTCGCAACCACTG 2124
Db       2041 GCATGGGCTATGCTCGCACTCATTGGTGCTGGGCAAGCTGAGGTAGACCGGAAACCACTA 2100
Qy       2125 CACCGGGCAGCTGCATACTTGATAAATTCCCAATTGGAGGACGGTGACTTTCCGCAGCAG 2184
Db       2101 CACCGGGCTGCAAGATACTTGATTAATGCTCAAATGGAGAATGGTGATTTTCCACAACAG 2160
Qy       2185 GAAATAATGGGAGTCTTCAACAAGAATTGCATGATCACATACGCCGCATACAGAAACATA 2244
Db       2161 GAAATAATGGGAGTCTTCAATAGGAACTGCATGATAACATATGCCGCGTATCGAAACATT 2220
Qy       2245 TTCCCCATTTGGGCGTTGGGAGAATACCAATCCCAAGTATT 2285
Db       2221 TTTCCGATATGGGCTTTGGGGGAGTACCGTTGTCAGGTATT 2261
```

RESULT 4
US-10-437-963-71137
; Sequence 71137, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 71137
; LENGTH: 4120
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_7163C.1
US-10-437-963-71137

Query Match 47.7%; Score 1222; DB 20; Length 4120;
Best Local Similarity 71.6%; Pred. No. 0;
Matches 1603; Conservative 0; Mismatches 635; Indels 0; Gaps 0;

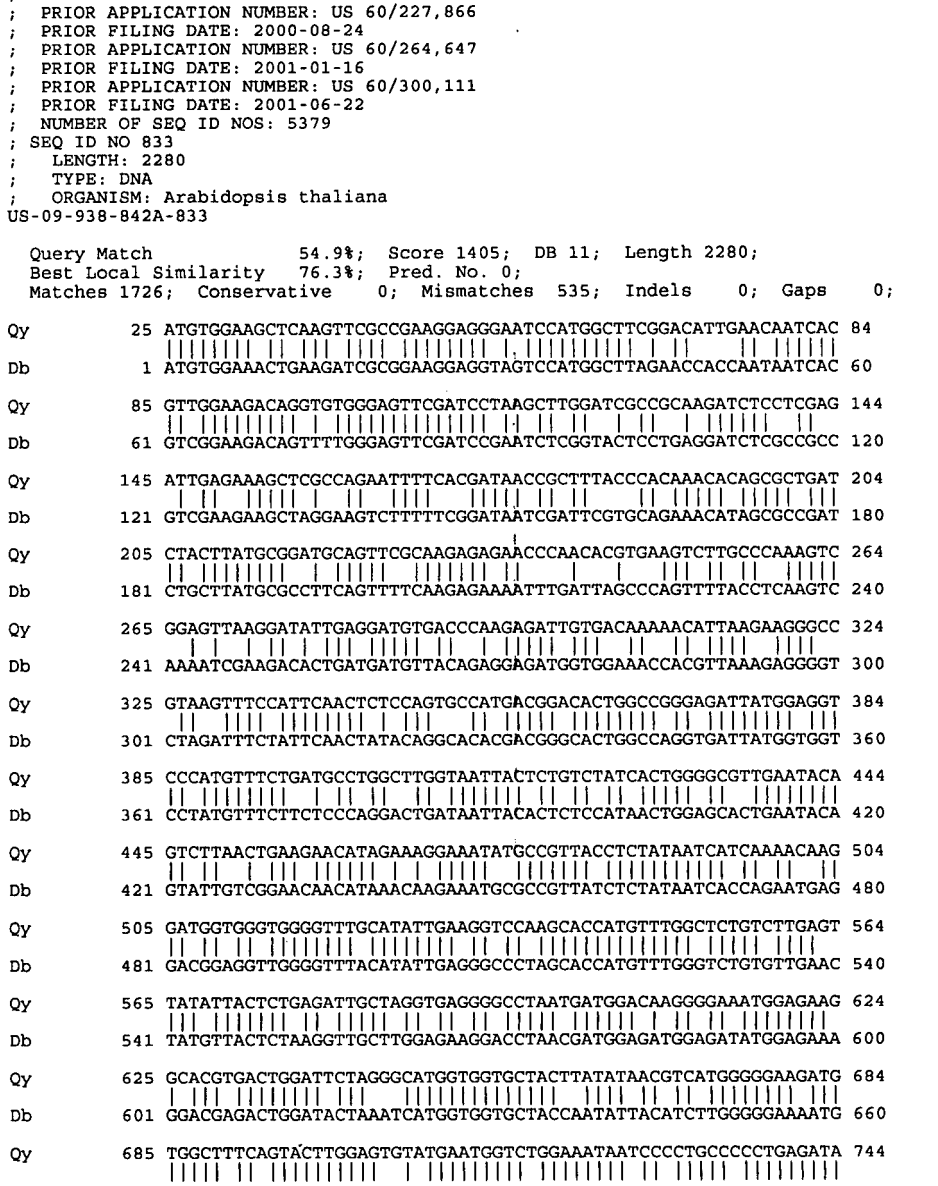
```
Qy         49 GGAGGGAATCCATGGCTTCGGACATTGAACAATCACGTTGGAAGACAGGTGTGGGAGTTC 108
Db         23 GGGGGGGACCCGTGGCTGCGCACGAAAAACGGCCACGTGGGGCGCCAGGTCTGGGAGTTC 82
Qy        109 GATCCTAAGCTTGGATCGCCGCAAGATCTCCTCGAGATTGAGAAAGCTCGCCAGAATTTT 168
Db         83 GACCCCGCCGCGGGCGACCCCGACGAGCTCGCCGCCGTCGAGGCCGCCCGCCGCGGTTTC 142
```



RESULT 3
US-09-938-842A-833
; Sequence 833, Application US/09938842A
; Publication No. US20040009476A9

; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIP1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 833
; LENGTH: 2280
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-833

Query Match 54.9%; Score 1405; DB 11; Length 2280;
Best Local Similarity 76.3%; Pred. No. 0;
Matches 1726; Conservative 0; Mismatches 535; Indels 0; Gaps 0;



```
Db       2101 GCTAAGAGAGATTCGCAACCACTGCACCGGGCAGCTGCATACTTGATAAATTCCCAATTG 2160

Qy       2161 GAGGACGGTGACTTTCCGCAGCAGGAAATAATGGGAGTCTTCAACAAGAATTGCATGATC 2220

Db       2161 GAGGACGGTGACTTTCCGCAGCAGGAAATAATGGGAGTCTTCAACAAGAATTGCATGATC 2220

Qy       2221 ACATACGCCGCATACAGAAACATATTCCCCATTTGGGCGTTGGGAGAATACCAATCCCAA 2280

Db       2221 ACATACGCCGCATACAGAAACATATTCCCCATTTGGGCGTTGGGAGAATACCAATCCCAA 2280

Qy       2281 GTATTGCAATCTCGTTAATCGAGCCTTAGTTAGGGTGTCATCCTAATTATTTCGACCTGT 2340

Db       2281 GTATTGCAATCTCGTTAATCGAGCCTTAGTTAGGGTGTCATCCTAATTATTTCGACCTGT 2340

Qy       2341 GGACAGTAAAAGAAATAATAATAATAACCTATTTTTCTTTTATTTCCATGGCTCTCTTAA 2400

Db       2341 GGACAGTAAAAGAAATAATAATAATAACCTATTTTTCTTTTATTTCCATGGCTCTCTTAA 2400

Qy       2401 AATGTTTGTGACTAATGAGTTTAGTAGTCAGCTAAAAAAAAAAGCAAACACGTGGAGAAT 2460

Db       2401 AATGTTTGTGACTAATGAGTTTAGTAGTCAGCTAAAAAAAAAAGCAAACACGTGGAGAAT 2460

Qy       2461 GCCTGTAAGTTTTTTCTATTACTCATAGACGCCTCTCTTGCTTTCCCTGCAGCAGAGGAA 2520

Db       2461 GCCTGTAAGTTTTTTCTATTACTCATAGACGCCTCTCTTGCTTTCCCTGCAGCAGAGGAA 2520

Qy       2521 TTAAATATATACATAAATAGAGATATAAAAAAAAAAAAAA 2560

Db       2521 TTAAATATATACATAAATAGAGATATAAAAAAAAAAAAAA 2560
```

RESULT 2
US-09-938-842A-833
; Sequence 833, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIP1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 833
; LENGTH: 2280
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-833

Query Match 54.9%; Score 1405; DB 9; Length 2280;
Best Local Similarity 76.3%; Pred. No. 0;
Matches 1726; Conservative 0; Mismatches 535; Indels 0; Gaps 0;

```
Qy         25 ATGTGGAAGCTCAAGTTCGCCGAAGGAGGGAATCCATGGCTTCGGACATTGAACAATCAC 84

Db          1 ATGTGGAAACTGAAGATCGCGGAAGGAGGTAGTCCATGGCTTAGAACCACCAATAATCAC 60

Qy         85 GTTGGAAGACAGGTGTGGGAGTTCGATCCTAAGCTTGGATCGCCGCAAGATCTCCTCGAG 144

Db         61 GTCGGAAGACAGTTTTGGGAGTTCGATCCGAATCTCGGTACTCCTGAGGATCTCGCCGCC 120

Qy        145 ATTGAGAAAGCTCGCCAGAATTTTCACGATAACCGCTTTACCCACAAACACAGCGCTGAT 204

Db        121 GTCGAAGAAGCTAGGAAGTCTTTTTCGGATAATCGATTCGTGCAGAAACATAGCGCCGAT 180

Qy        205 CTACTTATGCGGATGCAGTTCGCAAGAGAGAACCCAACACGTGAAGTCTTGCCCAAAGTC 264

Db        181 CTGCTTATGCGCCTTCAGTTTCAAGAGAAAATTTGATTAGCCCAGTTTTACCTCAAGTC 240

Qy        265 GGAGTTAAGGATATTGAGGATGTGACCCAAGAGATTGTGACAAAAACATTAAGAAGGGCC 324

Db        241 AAAATCGAAGACACTGATGATGTTACAGAGGAGATGGTGGAAACCACGTTAAAGAGGGGT 300

Qy        325 GTAAGTTTCCATTCAACTCTCCAGTGCCATGACGGACACTGGCCGGGAGATTATGGAGGT 384

Db        301 CTAGATTTCTATTCAACTATACAGGCACACGACGGGCACTGGCCAGGTGATTATGGTGGT 360

Qy        385 CCCATGTTTCTGATGCCTGGCTTGGTAATTACTCTGTCTATCACTGGGGCGTTGAATACA 444

Db        361 CCTATGTTTCTTCTCCCAGGACTGATAATTACACTCTCCATAACTGGAGCACTGAATACA 420

Qy        445 GTCTTAACTGAAGAACATAGAAAGGAAATATGCCGTTACCTCTATAATCATCAAAACAAG 504

Db        421 GTATTGTCGGAACAACATAAACAAGAAATGCGCCGTTATCTCTATAATCACCAGAATGAG 480

Qy        505 GATGGTGGGTGGGGTTTGCATATTGAAGGTCCAAGCACCATGTTTGGCTCTGTCTTGAGT 564

Db        481 GACGGAGGTTGGGGTTTACATATTGAGGGCCCTAGCACCATGTTTGGGTCTGTGTTGAAC 540

Qy        565 TATATTACTCTGAGATTGCTAGGTGAGGGGCCTAATGATGGACAAGGGGAAATGGAGAAG 624

Db        541 TATGTTACTCTAAGGTTGCTTGGAGAAGGACCTAACGATGGAGATGGAGATATGGAGAAA 600

Qy        625 GCACGTGACTGGATTCTAGGGCATGGTGGTGCTACTTATATAACGTCATGGGGGAAGATG 684

Db        601 GGACGAGACTGGATACTAAATCATGGTGGTGCTACCAATATTACATCTTGGGGGAAAATG 660

Qy        685 TGGCTTTCAGTACTTGGAGTGTATGAATGGTCTGGAAATAATCCCCTGCCCCCTGAGATA 744

Db        661 TGGCTATCGGTACTTGGAGCTTTTGAATGGTCCGGAAATAACCCACTGCCACCTGAGATA 720

Qy        745 TGGCTCCTTCCATACATGCTTCCATTTCATCCAGGAAGGATGTGGTGTCACTGCCGGATG 804

Db        721 TGGCTTCTCCCATATTTCCTGCCAATACATCCAGGAAGGATGTGGTGCCATTGTCGAATG 780

Qy        805 GTCTATTTGCCGATGTCCTACTTATATGGCAAGAGGTTTGTTGGTCCAATCTCACCAACA 864

Db        781 GTGTACTTGCCGATGTCGTATTTGTATGGAAAAAGGTTTGTGGGTCCCATAACGTCCACT 840

Qy        865 GTATTATCTTTGAGAAAAGAGCTTTATACAGTACCATACCATGATATAGATTGGGATCAG 924

Db        841 GTTTTATCACTGAGAAAGGAGCTTTTCACAGTACCATATCATGAAGTCAACTGGAATGAA 900

Qy        925 GCTCGCAATTTGTGTGCAAAGGAAGATTTGTACTATCCTCACCCACTTGTACAGGATATT 984

Db        901 GCACGCAACCTTTGCGCAAAGGAGGATTTATACTACCCACATCCACTTGTGCAAGATATT 960

Qy        985 CTTTGGGCATCTCTACACAAGTTCCTTGAGCCTATTCTGATGCATTGGCCTGGAAAAAGA 1044

Db        961 CTTTGGGCATCACTTCATAAGATTGTTGAGCCTGTTCTGATGCGATGGCCTGGTGCAAAT 1020

Qy       1045 TTGAGGGAAAAGGCTATTATTTCTGCATTGGAGCATATACATTACGAAGATGAGAATACT 1104

Db       1021 TTGAGAGAAAAGGCTATAAGAACCGCAATAGAACATATTCATTATGAAGATGAGAATACT 1080

Qy       1105 CGATATATTTGCATAGGTCCTGTAAATAAGGTGTTAAATATGCTTTGCTGTTGGGTGGAA 1164

Db       1081 AGGTACATCTGCATAGGTCCCGTGAACAAGGTATTAAATATGCTTTGCTGTTGGGTAGAA 1140

Qy       1165 GATCCAAATTCTGAGGCCTTCAAGTTGCATCTTCCCAGGATTTATGATTATCTATGGATT 1224

Db       1141 GACCCAAACTCAGAGGCTTTCAAGTTGCACCTACCAAGAATCCATGACTTTCTCTGGTTA 1200

Qy       1225 GCAGAAGATGGCATGAAAATGCAGGGCTACAATGGAAGTCAACTATGGGACACTGCTTTT 1284

Db       1201 GCTGAAGATGGAATGAAGATGCAGGGTTATAACGGAAGCCAGCTATGGGATACAGGTTTT 1260
```

Query Match 100.0%; Score 2560; DB 20; Length 2560;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2560; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 TTGGCCTCTTGCCAGCAAAACAGAATGTGGAAGCTCAAGTTCGCCGAAGGAGGGAATCCA 60
Db       1 TTGGCCTCTTGCCAGCAAAACAGAATGTGGAAGCTCAAGTTCGCCGAAGGAGGGAATCCA 60

Qy      61 TGGCTTCGGACATTGAACAATCACGTTGGAAGACAGGTGTGGGAGTTCGATCCTAAGCTT 120
Db      61 TGGCTTCGGACATTGAACAATCACGTTGGAAGACAGGTGTGGGAGTTCGATCCTAAGCTT 120

Qy     121 GGATCGCCGCAAGATCTCCTCGAGATTGAGAAAGCTCGCCAGAATTTTCACGATAACCGC 180
Db     121 GGATCGCCGCAAGATCTCCTCGAGATTGAGAAAGCTCGCCAGAATTTTCACGATAACCGC 180

Qy     181 TTTACCCACAAACACAGCGCTGATCTACTTATGCGGATGCAGTTCGCAAGAGAGAACCCA 240
Db     181 TTTACCCACAAACACAGCGCTGATCTACTTATGCGGATGCAGTTCGCAAGAGAGAACCCA 240

Qy     241 ACACGTGAAGTCTTGCCCAAAGTCGGAGTTAAGGATATTGAGGATGTGACCCAAGAGATT 300
Db     241 ACACGTGAAGTCTTGCCCAAAGTCGGAGTTAAGGATATTGAGGATGTGACCCAAGAGATT 300

Qy     301 GTGACAAAAACATTAAGAAGGGCCGTAAGTTTCCATTCAACTCTCCAGTGCCATGACGGA 360
Db     301 GTGACAAAAACATTAAGAAGGGCCGTAAGTTTCCATTCAACTCTCCAGTGCCATGACGGA 360

Qy     361 CACTGGCCGGGAGATTATGGAGGTCCCATGTTTCTGATGCCTGGCTTGGTAATTACTCTG 420
Db     361 CACTGGCCGGGAGATTATGGAGGTCCCATGTTTCTGATGCCTGGCTTGGTAATTACTCTG 420

Qy     421 TCTATCACTGGGGCGTTGAATACAGTCTTAACTGAAGAACATAGAAAGGAAATATGCCGT 480
Db     421 TCTATCACTGGGGCGTTGAATACAGTCTTAACTGAAGAACATAGAAAGGAAATATGCCGT 480

Qy     481 TACCTCTATAATCATCAAAACAAGGATGGTGGGTGGGGTTTGCATATTGAAGGTCCAAGC 540
Db     481 TACCTCTATAATCATCAAAACAAGGATGGTGGGTGGGGTTTGCATATTGAAGGTCCAAGC 540

Qy     541 ACCATGTTTGGCTCTGTCTTGAGTTATATTACTCTGAGATTGCTAGGTGAGGGGCCTAAT 600
Db     541 ACCATGTTTGGCTCTGTCTTGAGTTATATTACTCTGAGATTGCTAGGTGAGGGGCCTAAT 600

Qy     601 GATGGACAAGGGGAAATGGAGAAGGCACGTGACTGGATTCTAGGGCATGGTGGTGCTACT 660
Db     601 GATGGACAAGGGGAAATGGAGAAGGCACGTGACTGGATTCTAGGGCATGGTGGTGCTACT 660

Qy     661 TATATAACGTCATGGGGGAAGATGTGGCTTTCAGTACTTGGAGTGTATGAATGGTCTGGA 720
Db     661 TATATAACGTCATGGGGGAAGATGTGGCTTTCAGTACTTGGAGTGTATGAATGGTCTGGA 720

Qy     721 AATAATCCCCTGCCCCCTGAGATATGGCTCCTTCCATACATGCTTCCATTTCATCCAGGA 780
Db     721 AATAATCCCCTGCCCCCTGAGATATGGCTCCTTCCATACATGCTTCCATTTCATCCAGGA 780

Qy     781 AGGATGTGGTGTCACTGCCGGATGGTCTATTTGCCGATGTCCTACTTATATGGCAAGAGG 840
Db     781 AGGATGTGGTGTCACTGCCGGATGGTCTATTTGCCGATGTCCTACTTATATGGCAAGAGG 840

Qy     841 TTTGTTGGTCCAATCTCACCAACAGTATTATCTTTGAGAAAAGAGCTTTATACAGTACCA 900
Db     841 TTTGTTGGTCCAATCTCACCAACAGTATTATCTTTGAGAAAAGAGCTTTATACAGTACCA 900

Qy     901 TACCATGATATAGATTGGGATCAGGCTCGCAATTTGTGTGCAAAGGAAGATTTGTACTAT 960
Db     901 TACCATGATATAGATTGGGATCAGGCTCGCAATTTGTGTGCAAAGGAAGATTTGTACTAT 960

Qy     961 CCTCACCCACTTGTACAGGATATTCTTTGGGCATCTCTACACAAGTTCCTTGAGCCTATT 1020
Db     961 CCTCACCCACTTGTACAGGATATTCTTTGGGCATCTCTACACAAGTTCCTTGAGCCTATT 1020

Qy    1021 CTGATGCATTGGCCTGGAAAAAGATTGAGGGAAAAGGCTATTATTTCTGCATTGGAGCAT 1080
Db    1021 CTGATGCATTGGCCTGGAAAAAGATTGAGGGAAAAGGCTATTATTTCTGCATTGGAGCAT 1080

Qy    1081 ATACATTACGAAGATGAGAATACTCGATATATTTGCATAGGTCCTGTAAATAAGGTGTTA 1140
Db    1081 ATACATTACGAAGATGAGAATACTCGATATATTTGCATAGGTCCTGTAAATAAGGTGTTA 1140

Qy    1141 AATATGCTTTGCTGTTGGGTGGAAGATCCAAATTCTGAGGCCTTCAAGTTGCATCTTCCC 1200
Db    1141 AATATGCTTTGCTGTTGGGTGGAAGATCCAAATTCTGAGGCCTTCAAGTTGCATCTTCCC 1200

Qy    1201 AGGATTTATGATTATCTATGGATTGCAGAAGATGGCATGAAAATGCAGGGCTACAATGGA 1260
Db    1201 AGGATTTATGATTATCTATGGATTGCAGAAGATGGCATGAAAATGCAGGGCTACAATGGA 1260

Qy    1261 AGTCAACTATGGGACACTGCTTTTGCTGTCCAAGCAATTATTGCATCTAACCTCATTGAA 1320
Db    1261 AGTCAACTATGGGACACTGCTTTTGCTGTCCAAGCAATTATTGCATCTAACCTCATTGAA 1320

Qy    1321 GAATTTGGTCCAACTATAAGAAAAGCTCATACCTATATTAAGAATTCACAGGTTTTAGAA 1380
Db    1321 GAATTTGGTCCAACTATAAGAAAAGCTCATACCTATATTAAGAATTCACAGGTTTTAGAA 1380

Qy    1381 GATTGTCCAGGTGATCTTAATAAATGGTACCGTCACATTTCAAAAGGTGCTTGGCCTTTT 1440
Db    1381 GATTGTCCAGGTGATCTTAATAAATGGTACCGTCACATTTCAAAAGGTGCTTGGCCTTTT 1440

Qy    1441 TCAACTGGAGATCATGGATGGCCAATTTCTGACTGCACAGCTGAAGGACTGAAAGCTGTT 1500
Db    1441 TCAACTGGAGATCATGGATGGCCAATTTCTGACTGCACAGCTGAAGGACTGAAAGCTGTT 1500

Qy    1501 CTATTACTATCCAAAATTGCACCAGAAATAGTTGGTGAGCCAATAGACGTGAAGCGATTA 1560
Db    1501 CTATTACTATCCAAAATTGCACCAGAAATAGTTGGTGAGCCAATAGACGTGAAGCGATTA 1560

Qy    1561 TATGATTCTGTAAATGTCATTCTCTCACTACAGAATGAAGATGGTGGTTTTGCAACATAT 1620
Db    1561 TATGATTCTGTAAATGTCATTCTCTCACTACAGAATGAAGATGGTGGTTTTGCAACATAT 1620

Qy    1621 GAGCTTAAACGATCTTATAATTGGTTGGAGATAATCAATCCTGCTGAAACTTTTGGTGAC 1680
Db    1621 GAGCTTAAACGATCTTATAATTGGTTGGAGATAATCAATCCTGCTGAAACTTTTGGTGAC 1680

Qy    1681 ATCGTTATTGATTATCCTTATGTGGAATGTACATCAGCAGCGATTCAAGCTTTGGCATCA 1740
Db    1681 ATCGTTATTGATTATCCTTATGTGGAATGTACATCAGCAGCGATTCAAGCTTTGGCATCA 1740

Qy    1741 TTTAGGAAATTATATCCTGGGCATCGCCGAGAAGAAATACAACATTGTATCGATAAAGCC 1800
Db    1741 TTTAGGAAATTATATCCTGGGCATCGCCGAGAAGAAATACAACATTGTATCGATAAAGCC 1800

Qy    1801 ACTACCTTCATTGAAAAAATACAAGCTTCAGATGGATCATGGTATGGTTCTTGGGGAGTT 1860
Db    1801 ACTACCTTCATTGAAAAAATACAAGCTTCAGATGGATCATGGTATGGTTCTTGGGGAGTT 1860

Qy    1861 TGCTTCACTTACGGTGCTTGGTTTGGGGTAAAAGGTCTGATTGCTGCTGGAAGGAGTTTC 1920
Db    1861 TGCTTCACTTACGGTGCTTGGTTTGGGGTAAAAGGTCTGATTGCTGCTGGAAGGAGTTTC 1920

Qy    1921 AGTAATTGCTCAAGCATCCGTAAAGCTTGTGAATTTCTGCTGTCCAAGCAGCTTCCTTCT 1980
Db    1921 AGTAATTGCTCAAGCATCCGTAAAGCTTGTGAATTTCTGCTGTCCAAGCAGCTTCCTTCT 1980

Qy    1981 GGTGGCTGGGGAGAGAGTTATCTGTCCTGTCAAAACAAGGTGTATTCAAATCTGGAAGGC 2040
Db    1981 GGTGGCTGGGGAGAGAGTTATCTGTCCTGTCAAAACAAGGTGTATTCAAATCTGGAAGGC 2040

Qy    2041 AACAGGTCTCATGTGGTCAACACTGGGTGGGCTATGTTGGCTCTCATTGATGCTGGACAG 2100
Db    2041 AACAGGTCTCATGTGGTCAACACTGGGTGGGCTATGTTGGCTCTCATTGATGCTGGACAG 2100

Qy    2101 GCTAAGAGAGATTCGCAACCACTGCACCGGGCAGCTGCATACTTGATAAATTCCCAATTG 2160
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2005, 06:31:56 ; Search time 1979 Seconds
(without alignments)
10697.694 Million cell updates/sec

Title: US-09-419-679-5
Perfect score: 2560
Sequence: 1 ttggcctcttgccagcaaaa.....agatataaaaaaaaaaaaaa 2560

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9794790 seqs, 4134909567 residues

Total number of hits satisfying chosen parameters: 19589580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq:*
20: /cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq:*
21: /cgn2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq:*
22: /cgn2_6/ptodata/2/pubpna/US10I_PUBCOMB.seq:*
23: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
24: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq2:*
25: /cgn2_6/ptodata/2/pubpna/US11A_PUBCOMB.seq:*
26: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
27: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
28: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

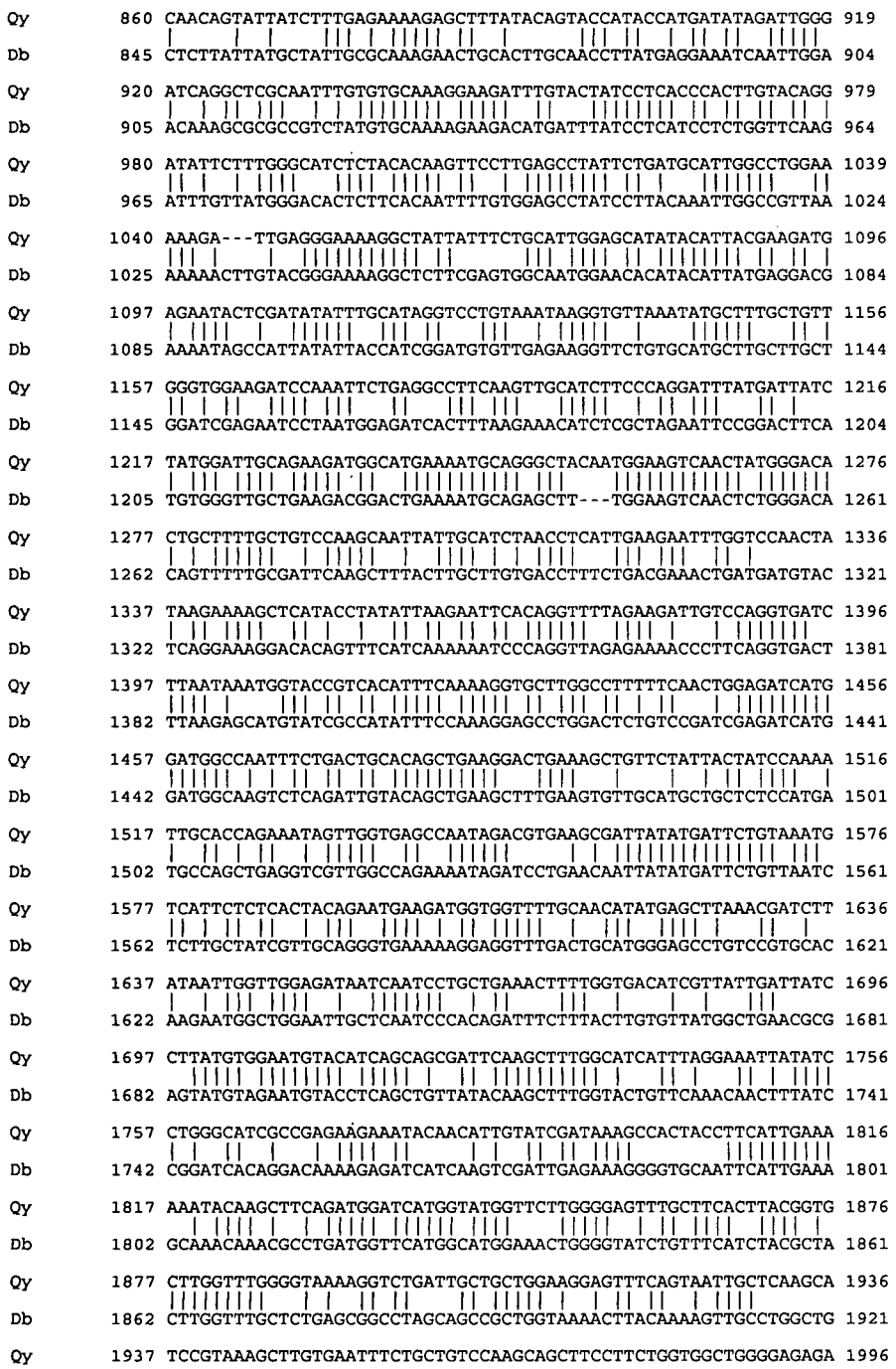
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2560	100.0	2560	20	US-10-607-726-5	Sequence 5, Appli
2	1405	54.9	2280	9	US-09-938-842A-833	Sequence 833, App
3	1405	54.9	2280	11	US-09-938-842A-833	Sequence 833, App
4	1222	47.7	4120	20	US-10-437-963-71137	Sequence 71137, A
5	1197.2	46.8	2776	21	US-10-425-115-91064	Sequence 91064, A
6	1196.2	46.7	2638	19	US-10-425-114-25096	Sequence 25096, A
7	1196.2	46.7	2786	21	US-10-425-115-90182	Sequence 90182, A
8	1183.6	46.2	2558	20	US-10-607-726-1	Sequence 1, Appli
9	1162	45.4	2355	20	US-10-437-963-71159	Sequence 71159, A
10	1145.6	44.7	2653	19	US-10-425-114-24593	Sequence 24593, A
11	944.2	36.9	2478	18	US-10-221-074-3	Sequence 3, Appli
12	944.2	36.9	2478	18	US-10-427-570A-4	Sequence 4, Appli
13	944.2	36.9	2669	19	US-10-424-599-35539	Sequence 35539, A
14	944	36.9	1882	20	US-10-607-726-3	Sequence 3, Appli
15	942.6	36.8	2567	19	US-10-425-114-7892	Sequence 7892, Ap
16	917.6	35.8	2593	19	US-10-425-114-24645	Sequence 24645, A
17	917.6	35.8	2627	21	US-10-425-115-65768	Sequence 65768, A
18	914.4	35.7	2574	21	US-10-739-930-2894	Sequence 2894, Ap
19	905	35.4	2731	18	US-10-221-074-1	Sequence 1, Appli
20	903.8	35.3	2538	18	US-10-221-074-7	Sequence 7, Appli
21	897.8	35.1	2746	19	US-10-424-599-31007	Sequence 31007, A
22	897.8	35.1	2766	18	US-10-221-074-5	Sequence 5, Appli
23	897.8	35.1	2766	18	US-10-427-570A-1	Sequence 1, Appli
24	847	33.1	2067	18	US-10-429-949-6	Sequence 6, Appli
25	843.8	33.0	2179	19	US-10-425-114-9157	Sequence 9157, Ap
26	819.6	32.0	2510	19	US-10-425-114-24639	Sequence 24639, A
27	817.8	31.9	2484	19	US-10-425-114-10317	Sequence 10317, A
28	817.8	31.9	2500	19	US-10-424-599-4661	Sequence 4661, Ap
29	810	31.6	2526	21	US-10-425-115-44214	Sequence 44214, A
30	804.2	31.4	2416	19	US-10-424-599-35538	Sequence 35538, A
31	796.2	31.1	2182	19	US-10-424-599-139406	Sequence 139406,
32	793	31.0	2093	18	US-10-259-194A-115	Sequence 115, App
33	787.8	30.8	1939	19	US-10-425-114-10363	Sequence 10363, A
34	769.8	30.1	2415	21	US-10-739-930-740	Sequence 740, App
35	766.4	29.9	2457	17	US-10-168-445-14	Sequence 14, Appl
36	756	29.5	761	19	US-10-425-114-10288	Sequence 10288, A
37	732	28.6	1546	18	US-10-259-194A-548	Sequence 548, App
38	722.2	28.2	2639	18	US-10-389-966-248	Sequence 248, App
39	716	28.0	1729	19	US-10-425-114-10855	Sequence 10855, A
40	695.8	27.2	1842	19	US-10-425-114-10336	Sequence 10336, A
41	677.4	26.5	1869	19	US-10-424-599-53604	Sequence 53604, A
42	658.2	25.7	1352	19	US-10-425-114-17202	Sequence 17202, A
43	655.8	25.6	2432	20	US-10-437-963-100852	Sequence 100852,
44	646.6	25.3	1862	20	US-10-437-963-63249	Sequence 63249, A
45	577	22.5	2414	21	US-10-425-115-171124	Sequence 171124,

ALIGNMENTS

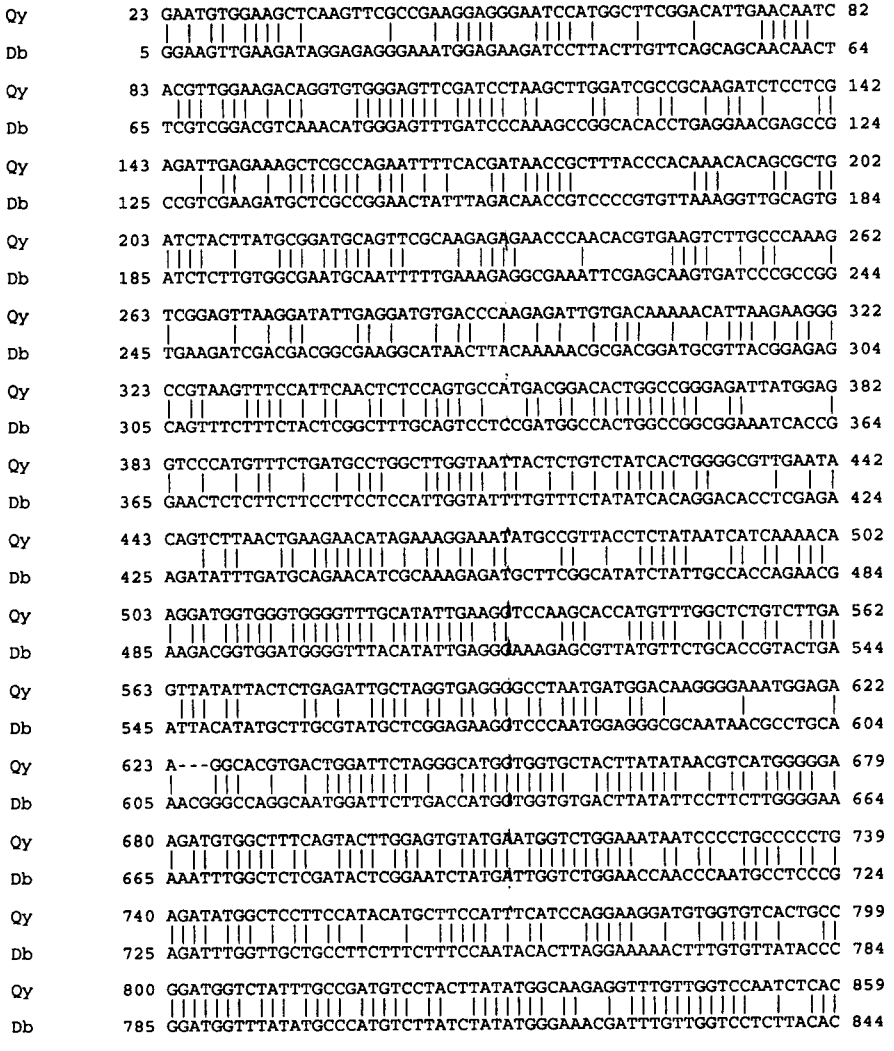
RESULT 1
US-10-607-726-5
; Sequence 5, Application US/10607726
; Publication No. US20040121439A1
; GENERAL INFORMATION:
; APPLICANT: Falco, S. Carl
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Hitz, William D.
; APPLICANT: Kinney, Anthony J.
; APPLICANT: Rafalski, Antoni
; APPLICANT: McGonigle, Brian
; APPLICANT: Lohman, Karin
; TITLE OF INVENTION: Enzymes Involved In Squalene Metabolizm
; FILE REFERENCE: BB1113 US NA
; CURRENT APPLICATION NUMBER: US/10/607,726
; CURRENT FILING DATE: 2003-06-27
; PRIOR APPLICATION NUMBER: US/09/419,679
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/105,405
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-10-23
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 5
; LENGTH: 2560
; TYPE: DNA
; ORGANISM: Glycine max
US-10-607-726-5



Search completed: November 5, 2005, 06:54:07
Job time : 1339 secs

PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 31.8%; Score 814.8; DB 3; Length 2532;
Best Local Similarity 60.7%; Pred. No. 2e-211;
Matches 1386; Conservative 0; Mismatches 887; Indels 9; Gaps 3;



```
Qy       1886 GGGTAAAAGGTCTGATTGCTGCTGGAAGGAGTTTCAGTAATTGCTCAAGCATCCGTAAAG 1945
Db       1646 CTCTTGGTGGTTTAGCAGCTGCTGGCAAGACTTATACCAATTGCGCTGCTATTCGCAAAG 1705

Qy       1946 CTTGTGAATTTCTGCTGTCCAAGCAGCTTCCTTCTGGTGGCTGGGGAGAGAGTTATCTGT 2005
Db       1706 CTGTTAAATTTCTTCTCACAACACAGAGAGAGGATGGTGGGTGGGGGGAGAGCTATCTTT 1765

Qy       2006 CCTGTCAAAACAAGGTGTATTCAAATCTGGAAGGCAACAGGTCTCATGTGGTCAACACTG 2065
Db       1766 CAAGCCCAAAAAAGATATATGTACCTCTCGAAGGAAGCCGATCCAATGTTGTACATACTG 1825

Qy       2066 GGTGGGCTATGTTGGCTCTCATTGATGCTGGACAGGCTAAGAGAGATTCGCAACCACTGC 2125
Db       1826 CATGGGCTCTTATGGGTTTAATTCATGCCGGCCAGGCAGAGAGAGACCCTACTCCTCTCC 1885

Qy       2126 ACCGGGCAGCTGCATACTTGATAAATTCCCAATTGGAGGACGGTGACTTTCCGCAGCAGG 2185
Db       1886 ATCGTGCTGCAAAATTGCTCATCAATTCCCAGTTGGAAGAAGGCGATTGGCCCCAACAGG 1945

Qy       2186 AAATAATGGGAGTCTTCAACAAGAATTGCATGATCACATACGCCGCATACAGAAACATAT 2245
Db       1946 AAATCACAGGAGTATTCATGAAAAATTGTATGTTGCATTACCCAATGTATAGAGATATTT 2005

Qy       2246 TCCCCATTTGGGCGTTGGGAGAATACC 2272
Db       2006 ACCCCTTGTGGGCTCTAGCCGAGTATC 2032
```

RESULT 15
AAC46759
ID AAC46759 standard; DNA; 2532 BP.
XX
AC AAC46759;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 51296.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.

XX
PR 04-MAY-2002; 2002US-0380159P.
XX
PA (ROBE-) ROBERTS NOBLE FOUND INC SAMMUEL.
XX
PI Dixon RA, Achnine L, Suzuki H, He X, Wang L;
XX
DR WPI; 2003-903669/82.
DR P-PSDB; ADF89799.
XX
PT Identifying a triterpene biosynthesis coding sequence, useful for crop
PT improvement, comprises contacting a cell from a target legume species
PT with methyl jasmonate and identifying a coding sequence which is
PT upregulated in the cell.
XX
PS Disclosure; SEQ ID NO 6; 181pp; English.
XX
CC The invention relates to identifying a triterpene biosynthesis coding
CC sequence. The mehtod involves obtaining a cell from a target legume
CC species, contacting the cell with methyl jasmonate, and identifying a
CC coding sequence which is specifically upregulated in the cell following
CC the contacting with methyl jasmonate to identify a triterpene
CC biosynthesis gene. Identifying a coding sequence further comprises
CC selecting a coding sequence having homology to a cytochrome P450,
CC glycosyltransferase, squalene synthase, squalene epoxidase or to a beta-
CC amyrin synthase. The method is useful in isolating plant genes and their
CC regulatory sequences involved in the biosynthesis of triterpene saponins.
CC The method allows genetic modification of the content and composition of
CC triterpene saponins in plants for crop improvement and in the development
CC of drugs, nutriceuticals and functional foods. The present sequence
CC represents a M. truncatula beta-amyrin synthase coding sequence.
XX
SQ Sequence 2067 BP; 577 A; 413 C; 496 G; 581 T; 0 U; 0 Other;

Query Match 33.1%; Score 847; DB 10; Length 2067;
Best Local Similarity 64.9%; Pred. No. 2.9e-220;
Matches 1303; Conservative 0; Mismatches 695; Indels 9; Gaps 3;

```
Qy   272 AGGATATTGAGGATGTGACCCAAGAGATTGTGACAAAAACATTAAGAAGGGCCGTAAGTT 331
Db    29 AAGATGGAGAAGAGATAACATATGAGAAAGCAACGACAACGTTGAGAAGGGGCACACACC 88

Qy   332 TCCATTCAACTCTCCAGTGCCATGACGGACACTGGCCGGGAGATTATGGAGGTCCCATGT 391
Db    89 ATCTAGCAGCATTGCAAACCAGTGATGGCCATTGGCCTGCTCAAATTGCAGGTCCTCTAT 148

Qy   392 TTCTGATGCCTGGCTTGGTAATTACTCTGTCTATCACTGGGGCGTTGAATACAGTCTTAA 451
Db   149 TTTTCATGCCTCCCTTGGTTTTCTGTGTCTACATTACTGGACATCTTGATTCCGTCTTCC 208

Qy   452 CTGAAGAACATAGAAAGGAAATATGCCGTTACCTCTATAATCATCAAAACAAGGATGGTG 511
Db   209 CACGAGAACATCGCAAAGAGATTCTTCGTTACATTTACTGTCACCAAAATGAAGATGGAG 268

Qy   512 GGTGGGGTTTGCATATTGAAGGTCCAAGCACCATGTTTGGCTCTGTCTTGAGTTATATTA 571
Db   269 GATGGGGGCTACACATTGAGGGTCACAGCACCATGTTTTGTACTGCACTTAACTATATAT 328

Qy   572 CTCTGAGATTGCTAGGTGAGGGGCCTAATGATGGACAAG---GGGAAATGGAGAAGGCAC 628
Db   329 GTATGCGAATTCTCGGAGAAGGACCTGATGGCGGTCAAGACAATGCTTGTGCTAGAGCCA 388

Qy   629 GTGACTGGATTCTAGGGCATGGTGGTGCTACTTATATAACGTCATGGGGGAAGATGTGGC 688
Db   389 GAAACTGGATTCGGGCACACGGTGGTGTCACATATATACCTTCATGGGGAAAAACTTGGC 448

Qy   689 TTTCAGTACTTGGAGTGTATGAATGGTCTGGAAATAATCCCCTGCCCCCTGAGATATGGC 748
Db   449 TTTCGATACTTGGTCTCTTTGATTGGTTGGGAAGCAACCCAATGCCCCCTGAGTTTGGA 508

Qy   749 TCCTTCCATACATGCTTCCATTTCATCCAGGAAGGATGTGGTGTCACTGCCGGATGGTCT 808
Db   509 TCCTTCCTTCATTTCTTCCTATGCATCCAGCTAAAATGTGGTGTTATTGTCGATTGGTAT 568

Qy   809 ATTTGCCGATGTCCTACTTATATGGCAAGAGGTTTGTTGGTCCAATCTCACCAACAGTAT 868
Db   569 ACATGCCTATGTCTTACTTGTACGGGAAGAGATTTGTGGGTCCGATCACACCACTCATCT 628

Qy   869 TATCTTTGAGAAAAGAGCTTTATACAGTACCATACCATGATATAGATTGGGATCAGGCTC 928
Db   629 TACAGTTGAGAGAAGAACTCCATACTCAGCCTTATGAAAAAATTAACTGGACGAAATCAC 688

Qy   929 GCAATTTGTGTGCAAAGGAAGATTTGTACTATCCTCACCCACTTGTACAGGATATTCTTT 988
Db   689 GTCACCTATGTGCAAAGGAAGATATTTACTATCCCCATCCTTTGATACAAGATCTGATAT 748

Qy   989 GGGCATCTCTACACAAGTTCCTTGAGCCTATTCTGATGCATTGGCCTGGAAAAAGATTG- 1047
Db   749 GGGATAGCTTATACATATTTACCGAGCCGCTTCTCACTCGCTGGCCTTTCAACAAGCTGG 808

Qy  1048 --AGGGAAAAGGCTATTATTTCTGCATTGGAGCATATACATTACGAAGATGAGAATACTC 1105
Db   809 TCAGAAAAAGAGCCCTTGAAGTTACAATGAAGCATATCCACTACGAGGATGAGAACAGTC 868

Qy  1106 GATATATTTGCATAGGTCCTGTAAATAAGGTGTTAAATATGCTTTGCTGTTGGGTGGAAG 1165
Db   869 GATACCTAACCATTGGGTGTGTGGAAAAGGTATTATGTATGCTTGCTTGTTGGGTGGAAG 928

Qy  1166 ATCCAAATTCTGAGGCCTTCAAGTTGCATCTTCCCAGGATTTATGATTATCTATGGATTG 1225
Db   929 ATCCAAATGGAGATGCTTACAAGAAGCATCTTGCAAGGGTCCAAGATTACTTGTGGATGT 988

Qy  1226 CAGAAGATGGCATGAAAATGCAGGGCTACAATGGAAGTCAACTATGGGACACTGCTTTTG 1285
Db   989 CAGAAGATGGAATGACCATGCAGAGTT---TTGGTAGCCAAGAATGGGATGCTGGTTTTG 1045

Qy  1286 CTGTCCAAGCAATTATTGCATCTAACCTCATTGAAGAATTTGGTCCAACTATAAGAAAAG 1345
Db  1046 CCGTTCAAGCTTTGCTTGCCGCTAACCTAAATGATGAAATCGAACCTGCACTTGCCAAAG 1105

Qy  1346 CTCATACCTATATTAAGAATTCACAGGTTTTAGAAGATTGTCCAGGTGATCTTAATAAAT 1405
Db  1106 GACATGATTTCATTAAGAAATCTCAGGTTACAGAGAACCCTTCTGGAGATTTTAAGAGTA 1165

Qy  1406 GGTACCGTCACATTTCAAAAGGTGCTTGGCCTTTTTCAACTGGAGATCATGGATGGCCAA 1465
Db  1166 TGCATCGTCATATTTCTAAAGGCTCATGGAGCTTCTCCGATCAAGACCATGGATGGCAAG 1225

Qy  1466 TTTCTGACTGCACAGCTGAAGGACTGAAAGCTGTTCTATTACTATCCAAAATTGCACCAG 1525
Db  1226 TTTCTGATTGCACCGCTGAAGGTTTGAAGTGTTGTCTACTTTTATCAATGTTGCCTCCAG 1285

Qy  1526 AAATAGTTGGTGAGCCAATAGACGTGAAGCGATTATATGATTCTGTAAATGTCATTCTCT 1585
Db  1286 AGATTGTGGGGGAAAAGATGGAACCAGAAAGGTTATATGATTCGGTCAATGTCTTGTTGT 1345

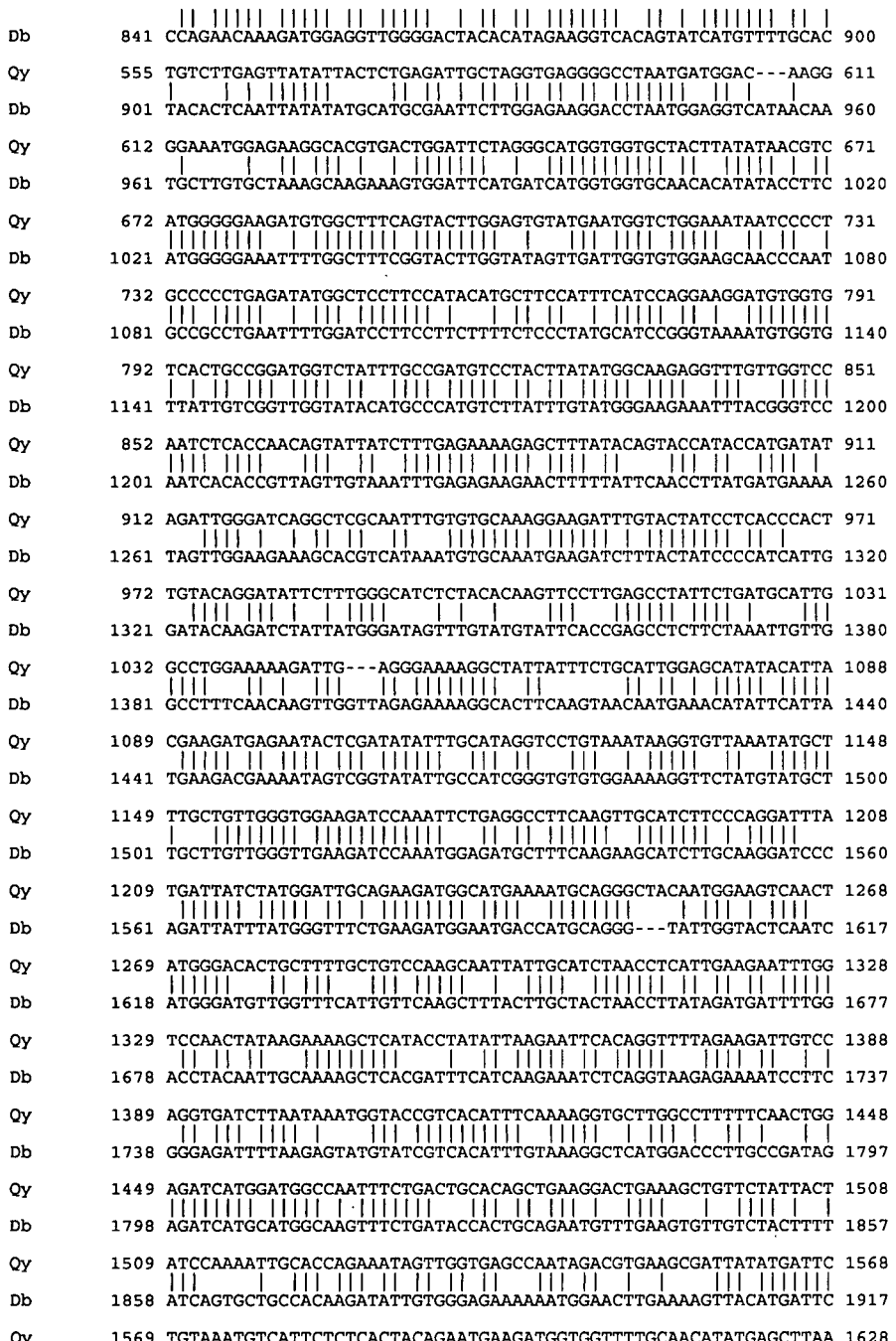
Qy  1586 CACTACAGAATGAAGATGGTGGTTTTGCAACATATGAGCTTAAACGATCTTATAATTGGT 1645
Db  1346 CGCTTCAGAGTAAAAAGGGTGGTTTGGCAGCATGGGAGCCCGCAGGAGCTCAAGAGTGGT 1405

Qy  1646 TGGAGATAATCAATCCTGCTGAAACTTTTGGTGACATCGTTATTGATTATCCTTATGTGG 1705
Db  1406 TAGAACTACTCAATCCCACTGAGTTTTTTGCGGACATTGTTGTTGAGCATGAATATGTTG 1465

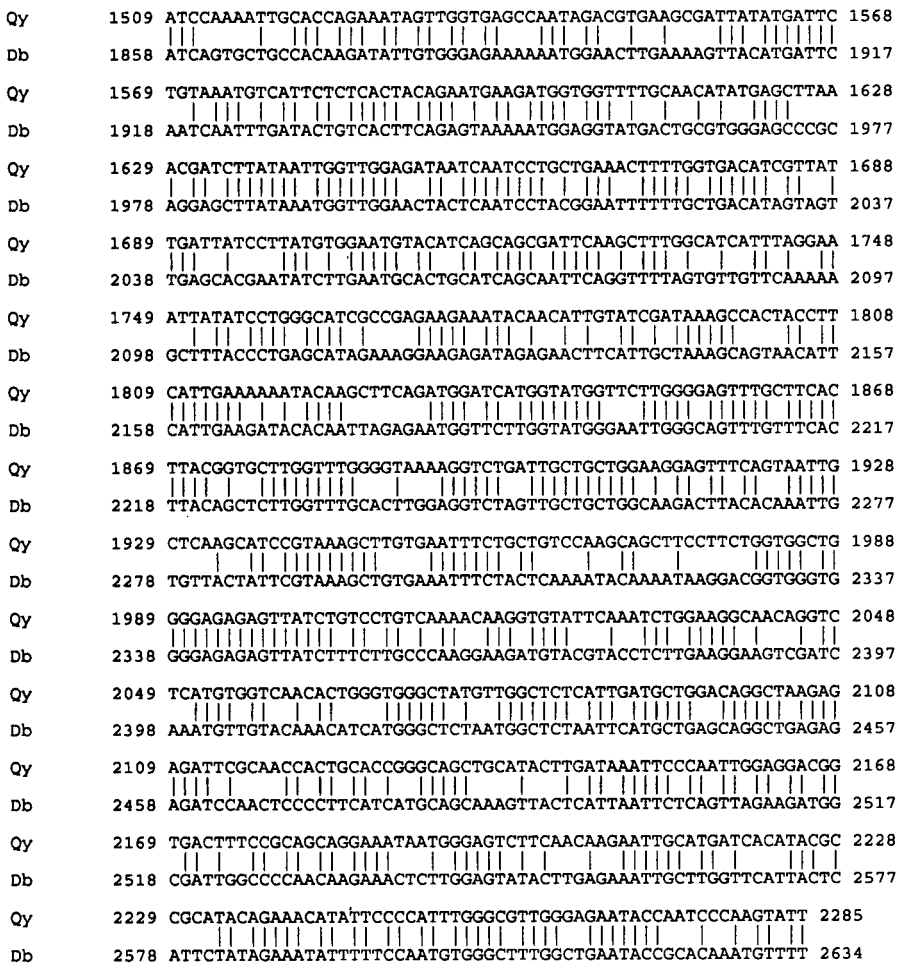
Qy  1706 AATGTACATCAGCAGCGATTCAAGCTTTGGCATCATTTAGGAAATTATATCCTGGGCATC 1765
Db  1466 AGTGCACTGGATCAGCAATTCAAGCTTTAGTTTTGTTCAAGAAGCTATATCCAGGGCATA 1525

Qy  1766 GCCGAGAAGAAATACAACATTGTATCGATAAAGCCACTACCTTCATTGAAAAAATACAAG 1825
Db  1526 GGAAGAAAGAGATAGAGAATTTCATCTCCGAGGCAGTTCGATTCATTGAAGATATACAAA 1585

Qy  1826 CTTCAGATGGATCATGGTATGGTTCTTGGGGAGTTTGCTTCACTTACGGTGCTTGGTTTG 1885
Db  1586 CAGCCGATGGTTCATGGTATGGAAACTGGGGAGTTTGCTTCACTTATGGTTCTTGGTTTG 1645
```

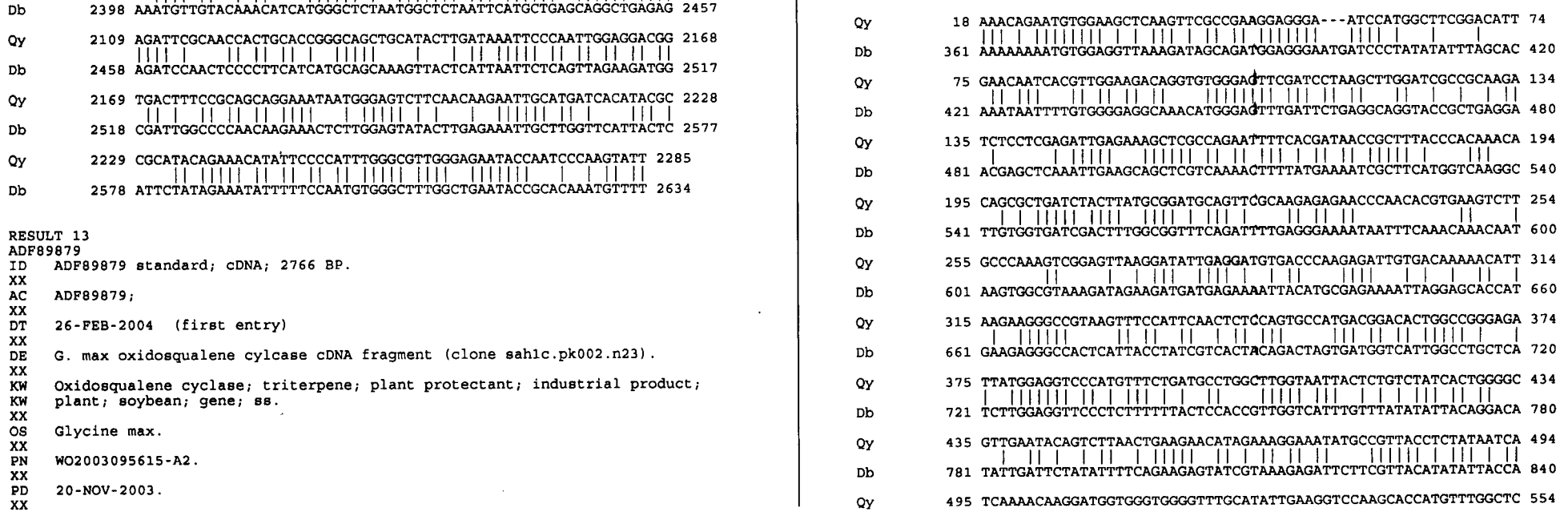


RESULT 14
ADF89798
ID ADF89798 standard; DNA; 2067 BP.
XX
AC ADF89798;
XX
DT 26-FEB-2004 (first entry)
XX
DE M. truncatula beta-amyrin synthase coding sequence.
XX
KW Triterpene; legume; methyl jasmonate; glycosyltransferase;
KW squalene synthase; squalene epoxidase; beta-amyrin synthase; plant;
KW saponin; genetic modification; drug development; nutriceutical; ds; gene.
XX
OS Medicago truncatula.
XX
FH Key Location/Qualifiers
FT CDS 1. .2067
FT /*tag= a
FT /product= " beta-amyrin synthase"
XX
PN WO2003093425-A2.
XX
PD 13-NOV-2003.
XX
PF 05-MAY-2003; 2003WO-US013618.



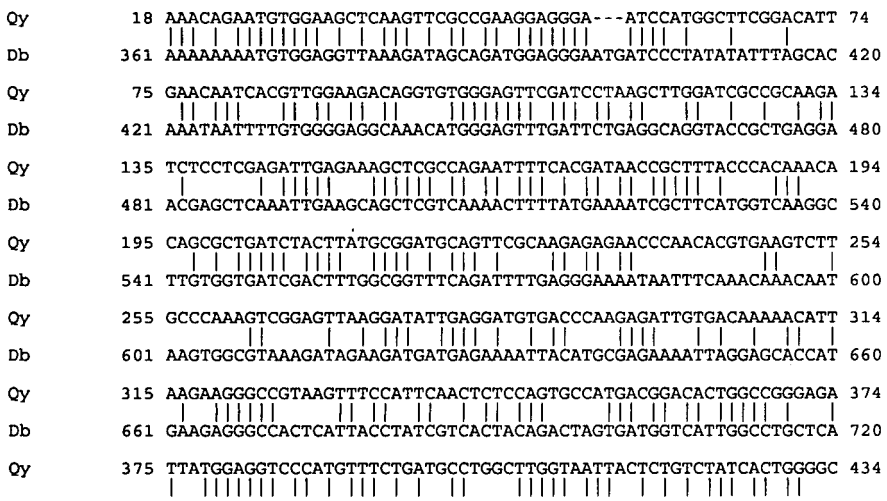
RESULT 13
ADF89879
ID ADF89879 standard; cDNA; 2766 BP.
XX
AC ADF89879;
XX
DT 26-FEB-2004 (first entry)
XX
DE G. max oxidosqualene cylcase cDNA fragment (clone sah1c.pk002.n23).
XX
KW Oxidosqualene cyclase; triterpene; plant protectant; industrial product;
KW plant; soybean; gene; ss.
XX
OS Glycine max.
XX
PN WO2003095615-A2.
XX
PD 20-NOV-2003.
XX
PF 05-MAY-2003; 2003WO-US014218.
XX
PR 09-MAY-2002; 2002US-0379361P.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Mcgonigle B, Maxwell CA, Hession AO;
XX
DR WPI; 2004-012098/01.
XX
PT New transgenic plants comprising recombinant DNA molecules encoding
PT oxidosqualene cyclase to lower triterpene level, used as feeds, foods,
PT beverages and industrial products like agricultural adjuvants or concrete
PT supplies.
XX
PS Example 1; SEQ ID NO 1; 39pp; English.
XX
CC The invention relates to a plant comprising at least one recombinant DNA
CC molecule comprising a promoter operably linked to at least a portion of
CC at least one oxidosqualene cyclase gene, the molecule sufficient to
CC suppress the production of a triterpene or any of its progeny, where the
CC progeny comprise the molecule. The oxidosqualene cyclase gene catalyzes
CC the cyclization of 2,3-oxidosqualene to form a triterpene, e.g. beta-
CC amyrin, lanosterol, lupeol, cycloartenol, alpha-amyrin, isomultiflorenol
CC or any combination. The plants are useful in generating products which
CC may be used as feeds, foods, beverages and industrial products like
CC agricultural adjuvants, concrete supplies, dielectric fluids, dust
CC suppressants, fuel additives, hydraulic fluids, industrial cleaners,
CC industrial lubricants, metalworking fluids, odor reduction, paint
CC strippers, printing inks, printing supplies or saw guide oils. Saponins
CC may play a defense role against pathogens in plant tissues. The present
CC sequence represents a cDNA insert from clone sah1c.pk002.n23 encoding a
CC G. max oxidosqualene cylcase.
XX
SQ Sequence 2766 BP; 858 A; 498 C; 586 G; 824 T; 0 U; 0 Other;

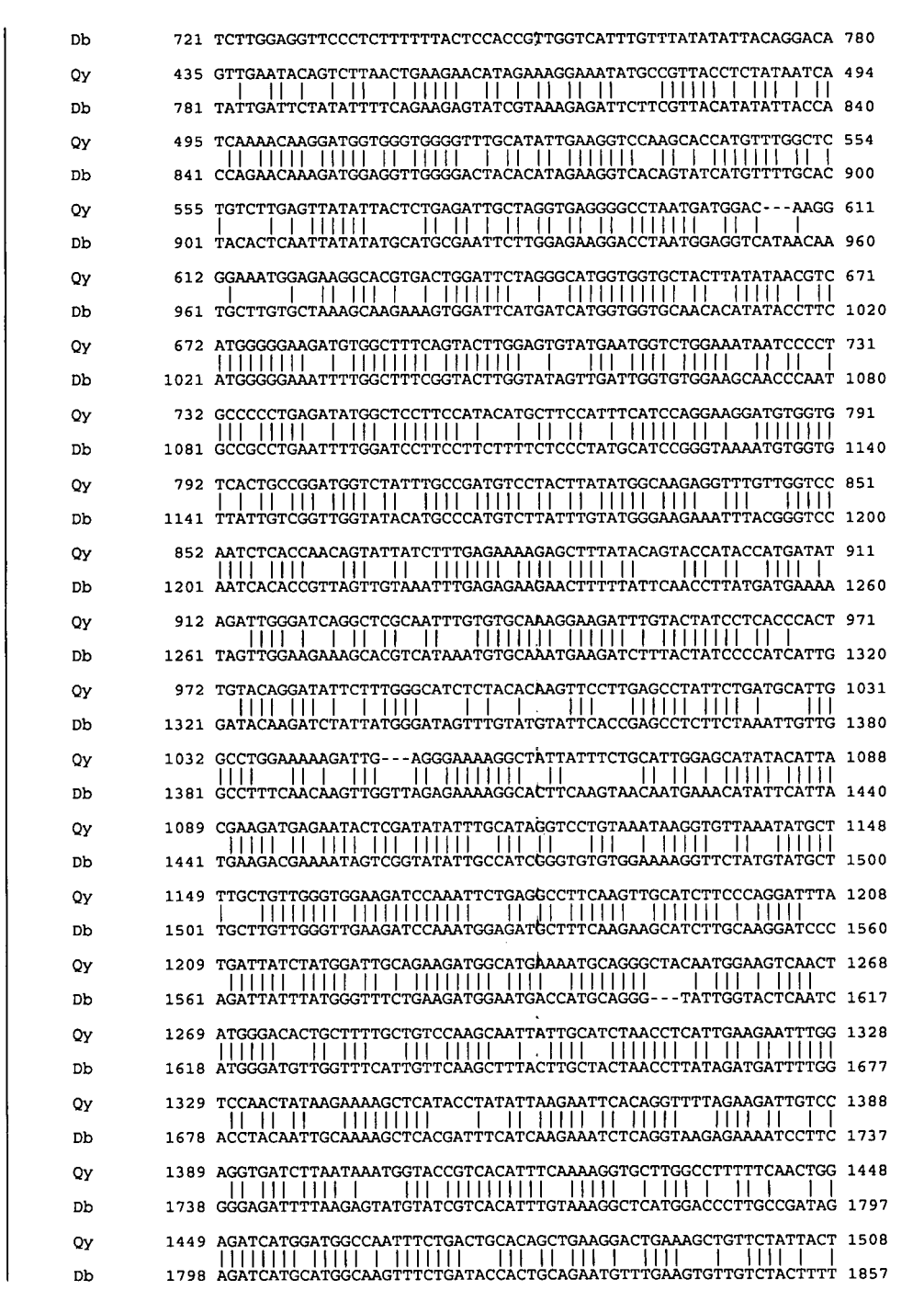
Query Match 35.1%; Score 897.8; DB 12; Length 2766;
Best Local Similarity 63.4%; Pred. No. 4.5e-234;
Matches 1443; Conservative 0; Mismatches 822; Indels 12; Gaps 4;

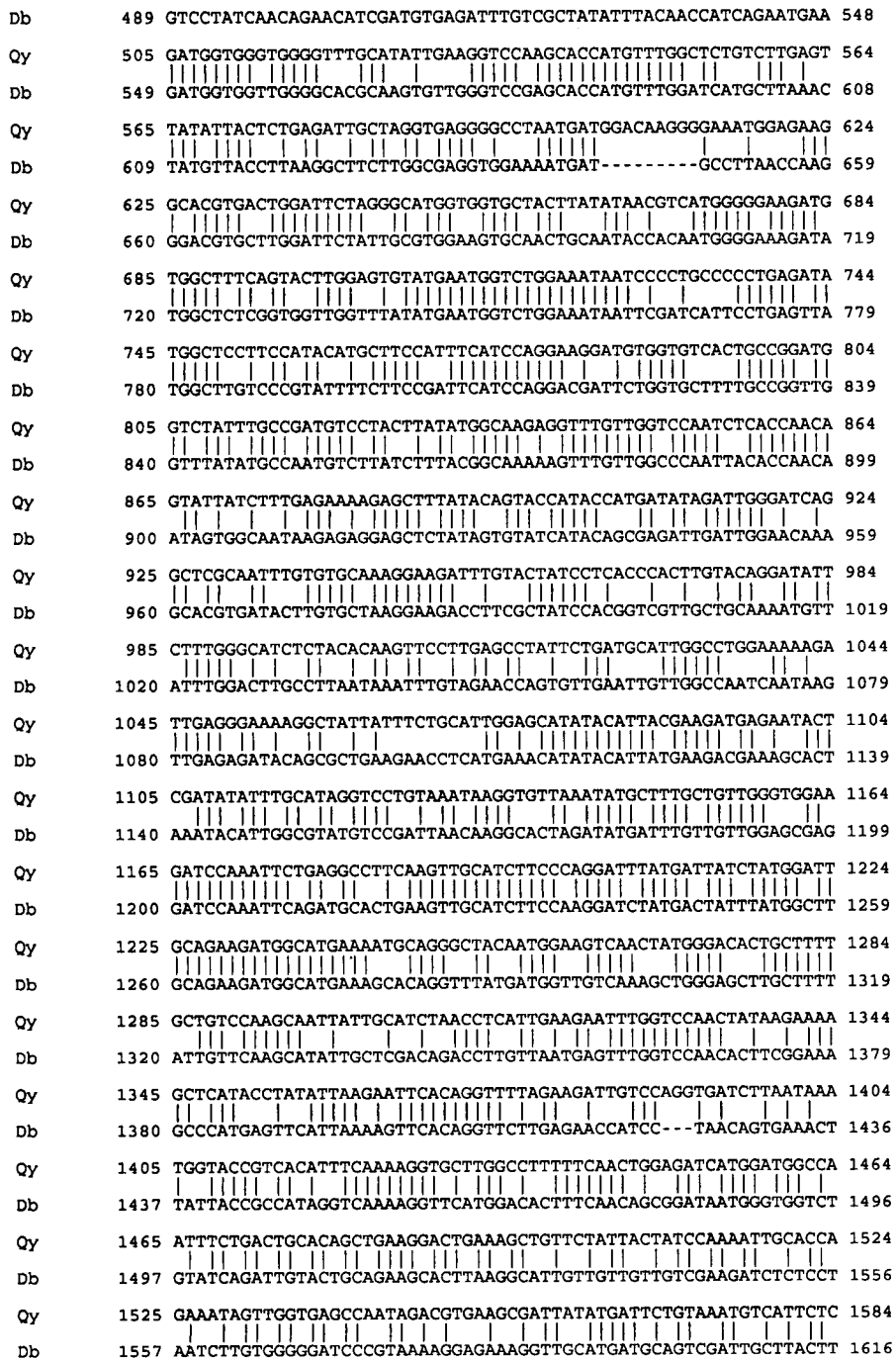


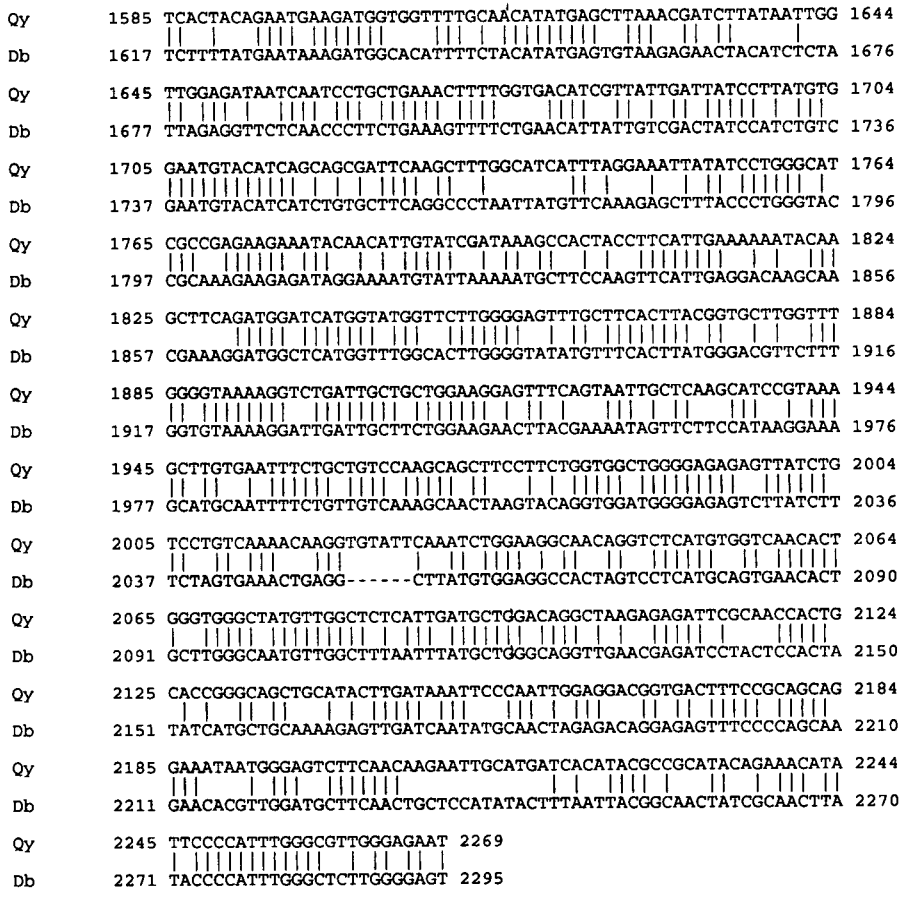
PD 13-SEP-2001.
XX
PF 09-MAR-2001; 2001WO-US007611.
XX
PR 09-MAR-2000; 2000US-0188054P.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Bryan GT, Mcgonigle B, Maxwell CA, Potter SM, Hwang D;
XX
DR WPI; 2001-565594/63.
DR P-PSDB; AAE09796.
XX
PT New nucleic acid fragment encoding oxidosqualene cyclase, useful as DNA
PT hybridization probe, polymerase chain reaction primer and for creating
PT transgenic plants with cells having higher or lower of oxidosqualene
PT cyclase.
XX
PS Claim 2; Page 47-48; 59pp; English.
XX
CC The invention relates to oxidosqualene cyclase proteins and their
CC corresponding cDNA molecules. Oxidosqualene cyclase genes are used as
CC hybridisation probes to screen libraries from any desired plant.
CC Oxidosqualene cyclase DNA are used to create transgenic plants with cells
CC having higher or lower level of oxidosqualene cyclase. These changes in
CC the plant seed are useful to improve the seed nutritional value, and in
CC the plant leaf may aid in insect tolerance. Oxidosqualene cyclase are
CC produced in heterologous host cells and are used to prepare antibodies.
CC The antibodies are useful for detecting oxidosqualene cyclase in situ in
CC cells or in vitro in cell extracts. The invention is useful for altering
CC level of expression of oxidosqualene cyclase (beta-amyrin synthase)
CC protein in a plant cell. Soybean grain produced from a transgenic plant
CC is useful for preparing food products such as soy protein product,
CC soybean meal, soy flour, soy protein concentrate, soy milk, a dietary
CC supplement, beans, nuts, tofu, natto, miso and tempeh. The formed soy
CC protein product having increased saponin levels is useful for lowering
CC cholesterol in a person suspected of having increased cholesterol, and
CC for decreasing risk of colon cancer in a person suspected of having risk
CC of cancer. The present cDNA sequence encodes soybean oxidosqualene
CC cyclase protein
XX
SQ Sequence 2766 BP; 858 A; 498 C; 586 G; 824 T; 0 U; 0 Other;

Query Match 35.1%; Score 897.8; DB 4; Length 2766;
Best Local Similarity 63.4%; Pred. No. 4.5e-234;
Matches 1443; Conservative 0; Mismatches 822; Indels 12; Gaps 4;

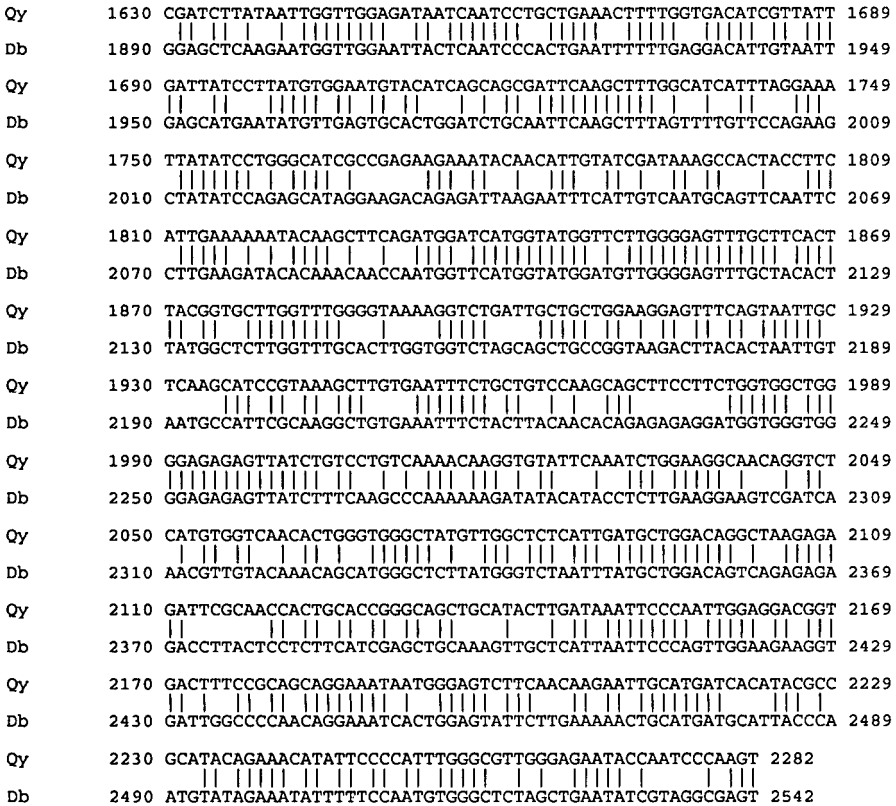






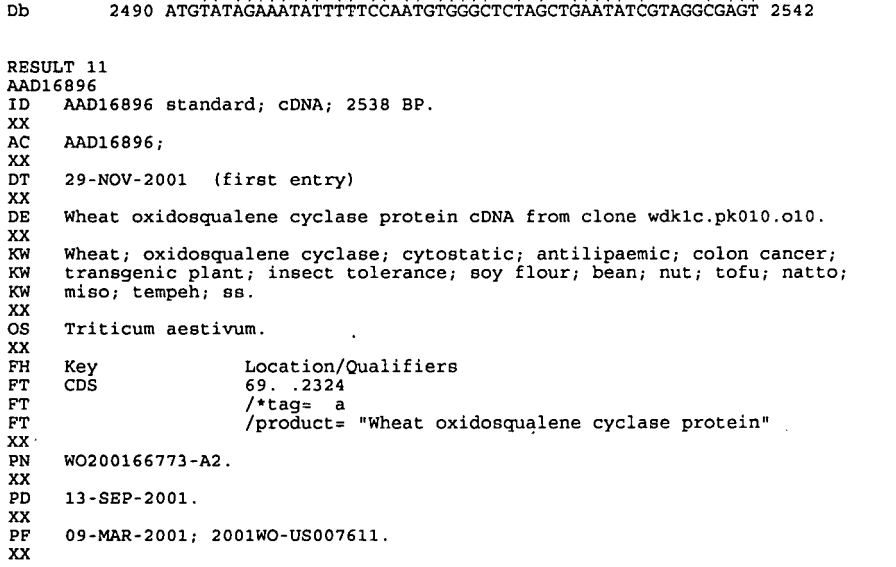


RESULT 12
AAD16895
ID AAD16895 standard; cDNA; 2766 BP.
XX
AC AAD16895;
XX
DT 29-NOV-2001 (first entry)
XX
DE Soybean oxidosqualene cyclase protein cDNA from clone sah1c.pk002.n23.
XX
KW Soybean; oxidosqualene cyclase; cytostatic; antilipaemic; colon cancer;
KW transgenic plant; insect tolerance; soy flour; bean; nut; tofu; natto;
KW miso; tempeh; ss.
XX
OS Glycine max.
XX
FH Key Location/Qualifiers
FT CDS 368. .2656
FT /*tag= a
FT /product= "Soybean oxidosqualene cyclase protein"
XX
PN WO200166773-A2.
XX



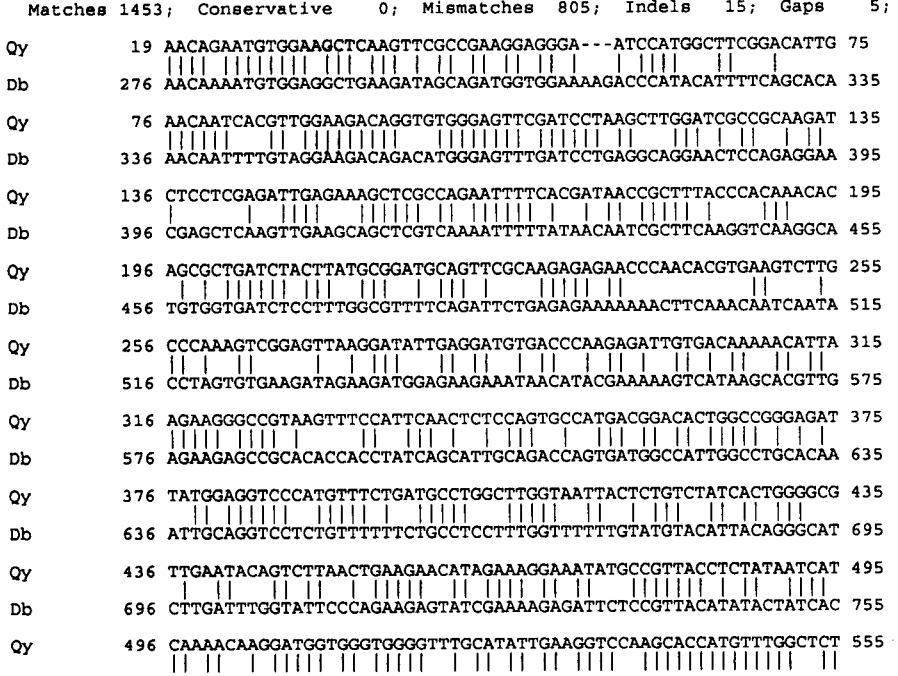
RESULT 11
AAD16896
ID AAD16896 standard; cDNA; 2538 BP.
XX
AC AAD16896;
XX
DT 29-NOV-2001 (first entry)
XX
DE Wheat oxidosqualene cyclase protein cDNA from clone wdk1c.pk010.o10.
XX
KW Wheat; oxidosqualene cyclase; cytostatic; antilipaemic; colon cancer;
KW transgenic plant; insect tolerance; soy flour; bean; nut; tofu; natto;
KW miso; tempeh; ss.
XX
OS Triticum aestivum.
XX
FH Key Location/Qualifiers
FT CDS 69. .2324
FT /*tag= a
FT /product= "Wheat oxidosqualene cyclase protein"
XX
PN WO200166773-A2.
XX
PD 13-SEP-2001.
XX
PF 09-MAR-2001; 2001WO-US007611.
XX
PR 09-MAR-2000; 2000US-0188054P.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Bryan GT, Mcgonigle B, Maxwell CA, Potter SM, Hwang D;
XX
DR WPI; 2001-565594/63.
DR P-PSDB; AAE09797.
XX
PT New nucleic acid fragment encoding oxidosqualene cyclase, useful as DNA
PT hybridization probe, polymerase chain reaction primer and for creating
PT transgenic plants with cells having higher or lower of oxidosqualene
PT cyclase.
XX
PS Claim 2; Page 50-51; 59pp; English.
XX
CC The invention relates to oxidosqualene cyclase proteins and their
CC corresponding cDNA molecules. Oxidosqualene cyclase genes are used as
CC hybridisation probes to screen libraries from any desired plant.
CC Oxidosqualene cyclase DNA are used to create transgenic plants with cells
CC having higher or lower level of oxidosqualene cyclase. These changes in
CC the plant seed are useful to improve the seed nutritional value, and in
CC the plant leaf may aid in insect tolerance. Oxidosqualene cyclase are
CC produced in heterologous host cells and are used to prepare antibodies.
CC The antibodies are useful for detecting oxidosqualene cyclase in situ in
CC cells or in vitro in cell extracts. The invention is useful for altering
CC level of expression of oxidosqualene cyclase (beta-amyrin synthase)
CC protein in a plant cell. Soybean grain produced from a transgenic plant
CC is useful for preparing food products such as soy protein product,
CC soybean meal, soy flour, soy protein concentrate, soy milk, a dietary
CC supplement, beans, nuts, tofu, natto, miso and tempeh. The formed soy
CC protein product having increased saponin levels is useful for lowering
CC cholesterol in a person suspected of having increased cholesterol, and
CC for decreasing risk of colon cancer in a person suspected of having risk
CC of cancer. The present cDNA sequence encodes wheat oxidosqualene cyclase
CC protein
XX
SQ Sequence 2538 BP; 727 A; 486 C; 613 G; 712 T; 0 U; 0 Other;

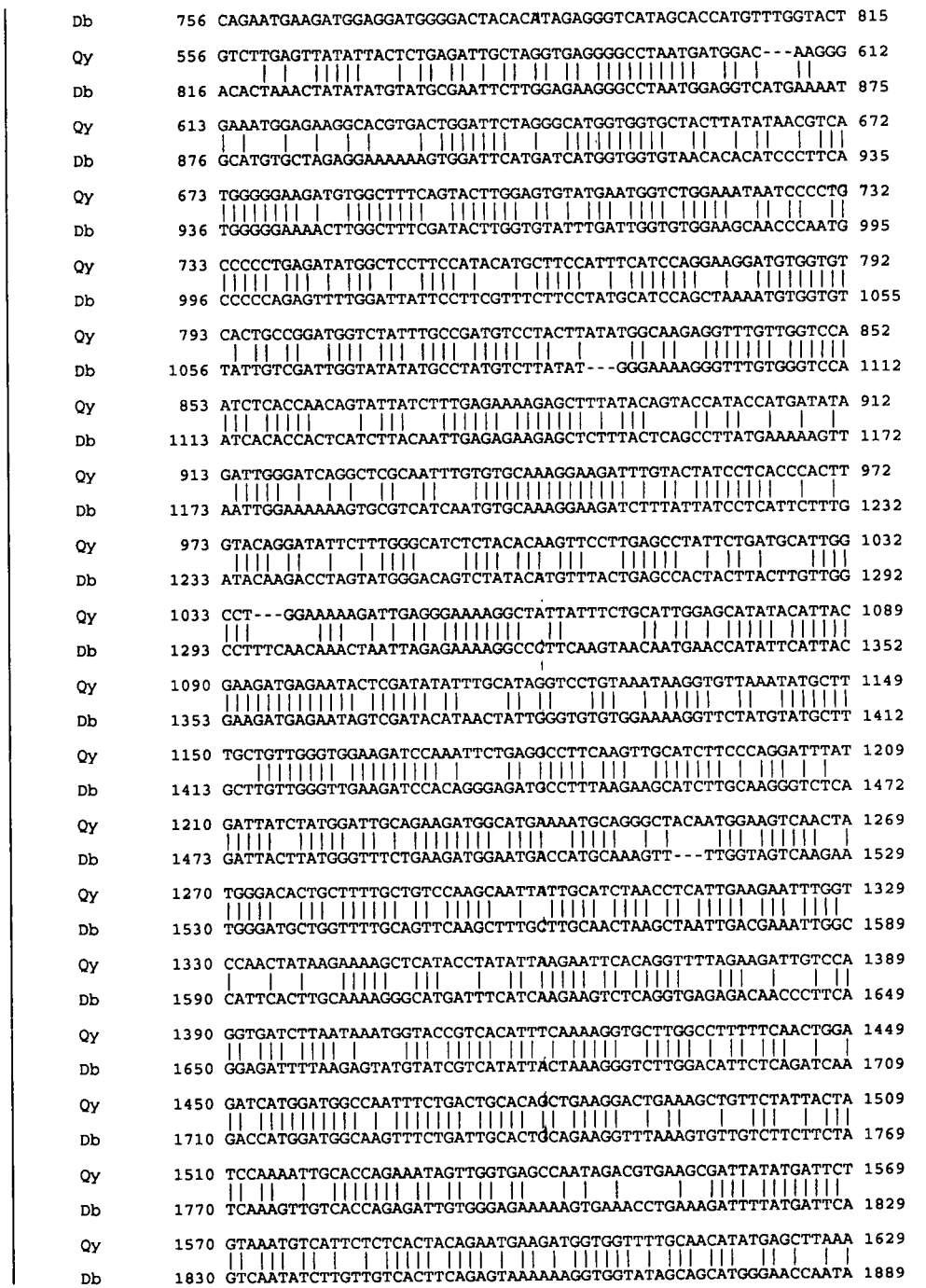
Query Match 35.3%; Score 903.8; DB 4; Length 2538;
Best Local Similarity 63.7%; Pred. No. 9.9e-236;
Matches 1430; Conservative 0; Mismatches 797; Indels 18; Gaps 3;



PI Bryan GT, Mcgonigle B, Maxwell CA, Potter SM, Hwang D;
XX
DR WPI; 2001-565594/63.
DR P-PSDB; AAE09794.
XX
PT New nucleic acid fragment encoding oxidosqualene cyclase, useful as DNA
PT hybridization probe, polymerase chain reaction primer and for creating
PT transgenic plants with cells having higher or lower of oxidosqualene
PT cyclase.
XX
PS Claim 2; Page 41-42; 59pp; English.
XX
CC The invention relates to oxidosqualene cyclase proteins and their
CC corresponding cDNA molecules. Oxidosqualene cyclase genes are used as
CC hybridisation probes to screen libraries from any desired plant.
CC Oxidosqualene cyclase DNA are used to create transgenic plants with cells
CC having higher or lower level of oxidosqualene cyclase. These changes in
CC the plant seed are useful to improve the seed nutritional value, and in
CC the plant leaf may aid in insect tolerance. Oxidosqualene cyclase are
CC produced in heterologous host cells and are used to prepare antibodies.
CC The antibodies are useful for detecting oxidosqualene cyclase in situ in
CC cells or in vitro in cell extracts. The invention is useful for altering
CC level of expression of oxidosqualene cyclase (beta-amyrin synthase)
CC protein in a plant cell. Soybean grain produced from a transgenic plant
CC is useful for preparing food products such as soy protein product,
CC soybean meal, soy flour, soy protein concentrate, soy milk, a dietary
CC supplement, beans, nuts, tofu, natto, miso and tempeh. The formed soy
CC protein product having increased saponin levels is useful for lowering
CC cholesterol in a person suspected of having increased cholesterol, and
CC for decreasing risk of colon cancer in a person suspected of having risk
CC of cancer. The present cDNA sequence encodes soybean oxidosqualene
CC cyclase protein
XX
SQ Sequence 2731 BP; 877 A; 458 C; 601 G; 795 T; 0 U; 0 Other;

Query Match 35.4%; Score 905; DB 4; Length 2731;
Best Local Similarity 63.9%; Pred. No. 4.8e-236;
Matches 1453; Conservative 0; Mismatches 805; Indels 15; Gaps 5;





SQ Sequence 1882 BP; 565 A; 363 C; 441 G; 513 T; 0 U; 0 Other;

Query Match 36.9%; Score 944; DB 12; Length 1882;
Best Local Similarity 74.4%; Pred. No. 9e-247;
Matches 1190; Conservative 0; Mismatches 410; Indels 0; Gaps 0;

```
Qy        687 GCTTTCAGTACTTGGAGTGTATGAATGGTCTGGAAATAATCCCCTGCCCCCTGAGATATG 746
              |||||| || ||||| || | ||| |||||||| || || ||  |||| || || |||||
Db         14 GCTTTCGGTGCTTGGTGTATTTGACTGGTCTGGCAACAACCCAGTGCCACCAGAAATATG 73

Qy        747 GCTCCTTCCATACATGCTTCCATTTCATCCAGGAAGGATGTGGTGTCACTGCCGGATGGT 806
              | |  | |||||  | || ||  |||||||||  | ||||||||||| ||||||||||
Db         74 GTTGTTGCCATATTTCCTGCCGATTCATCCAGGGCGAATGTGGTGTCATTGCCGGATGGT 133

Qy        807 CTATTTGCCGATGTCCTACTTATATGGCAAGAGGTTTGTTGGTCCAATCTCACCAACAGT 866
              |||||||| ||||  ||| | ||||| ||||||||||| || |||  |  ||||||   |
Db        134 TTATTTGCCTATGTGTTACATTTATGGAAAGAGGTTTGTGGGCCCAGTTACACCAATTAT 193

Qy        867 ATTATCTTTGAGAAAAGAGCTTTATACAGTACCATACCATGATATAGATTGGGATCAGGC 926
              |||     || ||||| || || ||   |||||| ||| ||||  | |||||||||  ||||
Db        194 ATTGGAATTAAGAAAGGAACTCTACGAAGTACCCTACAATGAAGTTGATTGGGACAAGGC 253

Qy        927 TCGCAATTTGTGTGCAAAGGAAGATTTGTACTATCCTCACCCACTTGTACAGGATATTCT 986
              ||||||| | ||||| ||||||||| ||||||||| || ||| | || ||||||  |  |
Db        254 TCGCAATCTATGTGCTAAGGAAGATCTGTACTATCCACATCCATTCGTGCAGGATGTATT 313

Qy        987 TTGGGCATCTCTACACAAGTTCCTTGAGCCTATTCTGATGCATTGGCCTGGAAAAAGATT 1046
              |||||  |||| ||||| ||  |||| ||   | || ||| |||||||| || | |||
Db        314 ATGGGCCACTCTCCACAAATTTGTTGAACCAGCTATGTTGCGTTGGCCTGGGAACAAATT 373

Qy       1047 GAGGGAAAAGGCTATTATTTCTGCATTGGAGCATATACATTACGAAGATGAGAATACTCG 1106
              |||||| || ||| |     |||   || ||||||| ||||  |||||||||| || ||
Db        374 GAGGGAGAAAGCTTTGGACACTGTCATGCAGCATATTCATTATGAAGATGAGAACACCCG 433

Qy       1107 ATATATTTGCATAGGTCCTGTAAATAAGGTGTTAAATATGCTTTGCTGTTGGGTGGAAGA 1166
              |||||||||||| ||||| ||||| ||||| ||||||||||||    || ||| | |||||
Db        434 ATATATTTGCATTGGTCCAGTAAACAAGGTATTAAATATGCTTGCTTGCTGGATTGAAGA 493

Qy       1167 TCCAAATTCTGAGGCCTTCAAGTTGCATCTTCCCAGGATTTATGATTATCTATGGATTGC 1226
              |||||| || ||||| |||||  | ||  |||| ||  |  | ||||| ||||||||||
Db        494 TCCAAACTCAGAGGCATTCAAACTCCACATTCCAAGAGTCCACGATTACCTATGGATTGC 553

Qy       1227 AGAAGATGGCATGAAAATGCAGGGCTACAATGGAAGTCAACTATGGGACACTGCTTTTGC 1286
              |||||||||||||||||||||||| || |||||||| || || |||||||| |||||  |
Db        554 AGAAGATGGCATGAAAATGCAGGGTTATAATGGAAGCCAGCTGTGGGACACAGCTTTCAC 613

Qy       1287 TGTCCAAGCAATTATTGCATCTAACCTCATTGAAGAATTTGGTCCAACTATAAGAAAAGC 1346
              || ||||| ||  | ||  ||  |||||||||||||||||||| ||| | | | |||
Db        614 AGTTCAAGCTATAGTGGCTACTGGCCTCATTGAAGAATTTGGTCCTACTCTTAAACTAGC 673

Qy       1347 TCATACCTATATTAAGAATTCACAGGTTTTAGAAGATTGTCCAGGTGATCTTAATAAATG 1406
              |||  ||| || |||||  | |||||| | || || || || || ||||||| | | ||
Db        674 ACATGGCTACATAAAGAAAACGCAGGTTATCGATGACTGCCCTGGAGATCTTAGTCAGTG 733

Qy       1407 GTACCGTCACATTTCAAAAGGTGCTTGGCCTTTTTCAACTGGAGATCATGGATGGCCAAT 1466
              |||||| ||||| || |||||||| ||||| ||||| ||||  |||||||| ||||| ||
Db        734 GTACCGCCACATATCTAAAGGTGCATGGCCCTTTTCTACTGCTGATCATGGTTGGCCTAT 793

Qy       1467 TTCTGACTGCACAGCTGAAGGACTGAAAGCTGTTCTATTACTATCCAAAATTGCACCAGA 1526
              || || ||||| || |||||||| || || |   |||| ||||| || ||| | |||||
Db        794 ATCAGATTGCACTGCAGAAGGACTTAAGGCGGCATTATTGCTATCGAAGATTTCTCCAGA 853

Qy       1527 AATAGTTGGTGAGCCAATAGACGTGAAGCGATTATATGATTCTGTAAATGTCATTCTCTC 1586
              || ||||| ||  ||·| || || ||  || | |||||||||| |||   |  | ||
Db        854 TATTGTTGGCGAAGCAGTGGAAGTTAATAGACTGTATGATTCTGTCAATTGTTTGATGTC 913

Qy       1587 ACTACAGAATGAAGATGGTGGTTTTGCAACATATGAGCTTAAACGATCTTATAATTGGTT 1646
              |      ||||||  |||||| |||||||||||||| || | | ||||||   ||| |
Db        914 ATACATGAATGATAATGGTGGATTTGCAACATATGAACTCACAAGGTCTTATGCCTGGCT 973

Qy       1647 GGAGATAATCAATCCTGCTGAAACTTTTGGTGACATCGTTATTGATTATCCTTATGTGGA 1706
              |||| | |||||||||||||| || ||||| ||||| || ||||||||||||||||||||
Db        974 GGAGCTTATCAATCCTGCTGAGACCTTTGGGGACATTGTGATTGATTATCCTTATGTGGA 1033

Qy       1707 ATGTACATCAGCAGCGATTCAAGCTTTGGCATCATTTAGGAAATTATATCCTGGGCATCG 1766
              ||| || |||||||| ||||| ||  || || ||||||  ||  | || ||||| || ||
Db        1034 ATGCACTTCAGCAGCAATTCAGGCCCTGACAGCATTTAAAAAGCTCTACCCTGGACACCG 1093

Qy       1767 CCGAGAAGAAATACAACATTGTATCGATAAAGCCACTACCTTCATTGAAAAAATACAAGC 1826
              |      |||||| |  | |||||     |||||  ||| ||| |||||     || |||
Db        1094 CAAGAGTGAAATAGACAACTGTATAAGCAAAGCTGCTAGCTTTATTGAGGGTATTCAAAA 1153

Qy       1827 TTCAGATGGATCATGGTATGGTTCTTGGGGAGTTTGCTTCACTTACGGTGCTTGGTTTGG 1886
                 ||||| ||||||||||||||||||  ||||| || || || ||  | ||||||||
Db        1154 AAGCGATGGTTCATGGTATGGTTCTTGGGCTGTTTGTTTTACCTATGGCACATGGTTTGG 1213

Qy       1887 GGTAAAAGGTCTGATTGCTGCTGGAAGGAGTTTCAGTAATTGCTCAAGCATCCGTAAAGC 1946
              ||||| ||  |  |||||||||| ||||  ||||  ||  |  |     ||| | || ||
Db        1214 TGTAAAGGGATTAGTTGCTGCTGGTAGGACATTCAAAAACAGTCCTGCAATCAGAAAGGC 1273

Qy       1947 TTGTGAATTTCTGCTGTCCAAGCAGCTTCCTTCTGGTGGCTGGGGAGAGAGTTATCTGTC 2006
              ||||| ||| || |||| ||  |||||||||||| |||||||||| || ||| ||||
Db        1274 ATGTGACTTTTTGTTGTCAAAAGAGCTTCCTTCTGGAGGCTGGGGAGAAAGCTATTTGTC 1333

Qy       2007 CTGTCAAAACAAGGTGTATTCAAATCTGGAAGGCAACAGGTCTCATGTGGTCAACACTGG 2066
              |  ||| |  |||| ||| | ||||| ||||| ||  |  |||||| ||| ||||||||
Db        1334 ATCCCAAGATCAGGTTTATACCAATCTCGAAGGGAAGCGACCTCATGCGGTGAACACTGG 1393

Qy       2067 GTGGGCTATGTTGGCTCTCATTGATGCTGGACAGGCTAAGAGAGATTCGCAACCACTGCA 2126
              ||||| ||| | || || || ||||| || |||||| |||||||| |    ||  ||||
Db        1394 TTGGGCCATGCTAGCCCTAATCGATGCAGGGCAGGCTGAGAGAGATCCAATTCCTTTGCA 1453

Qy       2127 CCGGGCAGCTGCATACTTGATAAATTCCCAATTGGAGGACGGTGACTTTCCGCAGCAGGA 2186
              || |||||          ||||| || |  |||| ||| || ||||| ||||| ||||| ||
Db        1454 TCGAGCAGCGAAGGTTTTGATCAACTTACAATCGGAAGATGGTGAATTTCCCCAGCAAGA 1513

Qy       2187 AATAATGGGAGTCTTCAACAAGAATTGCATGATCACATACGCCGCATACAGAAACATATT 2246
              || || |||||||||||||| || ||||||||||  ||| |||  || |||||||| ||
Db        1514 GATCATTGGAGTCTTCAACAAAAACTGCATGATCAGCTACTCCGAGTATAGAAACATCTT 1573

Qy       2247 CCCCATTTGGGCGTTGGGAGAATACCAATCCCAAGTATTG 2286
              ||| ||||||||  | || || ||||    ||  || |||
Db        1574 CCCTATTTGGGCCCTTGGGGAGTACCGTCGCCGCGTCTTG 1613
```

RESULT 10
AAD16893
ID AAD16893 standard; cDNA; 2731 BP.
XX
AC AAD16893;
XX
DT 29-NOV-2001 (first entry)
XX
DE Soybean oxidosqualene cyclase protein cDNA from clone sdp3c.pk020.o10.
XX
KW Soybean; oxidosqualene cyclase; cytostatic; antilipaemic; colon cancer;
KW transgenic plant; insect tolerance; soy flour; bean; nut; tofu; natto;
KW miso; tempeh; ss.
XX
OS Glycine max.
XX
FH Key Location/Qualifiers
FT CDS 282. .2567
FT /*tag= a
FT /product= "Soybean oxidosqualene cyclase protein"
XX
PN WO200166773-A2.
XX
PD 13-SEP-2001.
XX
PF 09-MAR-2001; 2001WO-US007611.
XX
PR 09-MAR-2000; 2000US-0188054P.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX

```
Qy       1227 AGAAGATGGCATGAAAATGCAGGGCTACAATGGAAGTCAACTATGGGACACTGCTTTTGC 1286
Db        554 AGAAGATGGCATGAAAATGCAGGGTTATAATGGAAGCCAGCTGTGGGACACAGCTTTCAC 613

Qy       1287 TGTCCAAGCAATTATTGCATCTAACCTCATTGAAGAATTTGGTCCAACTATAAGAAAAGC 1346
Db        614 AGTTCAAGCTATAGTGGCTACTGGCCTCATTGAAGAATTTGGTCCTACTCTTAAACTAGC 673

Qy       1347 TCATACCTATATTAAGAATTCACAGGTTTTAGAAGATTGTCCAGGTGATCTTAATAAATG 1406
Db        674 ACATGGCTACATAAAGAAAACGCAGGTTATCGATGACTGCCCTGGAGATCTTAGTCAGTG 733

Qy       1407 GTACCGTCACATTTCAAAAGGTGCTTGGCCTTTTTCAACTGGAGATCATGGATGGCCAAT 1466
Db        734 GTACCGCCACATATCTAAAGGTGCATGGCCCTTTTCTACTGCTGATCATGGTTGGCCTAT 793

Qy       1467 TTCTGACTGCACAGCTGAAGGACTGAAAGCTGTTCTATTACTATCCAAAATTGCACCAGA 1526
Db        794 ATCAGATTGCACTGCAGAAGGACTTAAGGCGGCATTATTGCTATCGAAGATTTCTCCAGA 853

Qy       1527 AATAGTTGGTGAGCCAATAGACGTGAAGCGATTATATGATTCTGTAAATGTCATTCTCTC 1586
Db        854 TATTGTTGGCGAAGCAGTGGAAGTTAATAGACTGTATGATTCTGTCAATTGTTTGATGTC 913

Qy       1587 ACTACAGAATGAAGATGGTGGTTTTGCAACATATGAGCTTAAACGATCTTATAATTGGTT 1646
Db        914 ATACATGAATGATAATGGTGGATTTGCAACATATGAACTCACAAGGTCTTATGCCTGGCT 973

Qy       1647 GGAGATAATCAATCCTGCTGAAACTTTTGGTGACATCGTTATTGATTATCCTTATGTGGA 1706
Db        974 GGAGCTTATCAATCCTGCTGAGACCTTTGGGGACATTGTGATTGATTATCCTTATGTGGA 1033

Qy       1707 ATGTACATCAGCAGCGATTCAAGCTTTGGCATCATTTAGGAAATTATATCCTGGGCATCG 1766
Db        1034 ATGCACTTCAGCAGCAATTCAGGCCCTGACAGCATTTAAAAAGCTCTACCCTGGACACCG 1093

Qy       1767 CCGAGAAGAAATACAACATTGTATCGATAAAGCCACTACCTTCATTGAAAAAATACAAGC 1826
Db        1094 CAAGAGTGAAATAGACAACTGTATAAGCAAAGCTGCTAGCTTTATTGAGGGTATTCAAAA 1153

Qy       1827 TTCAGATGGATCATGGTATGGTTCTTGGGGAGTTTGCTTCACTTACGGTGCTTGGTTTGG 1886
Db        1154 AAGCGATGGTTCATGGTATGGTTCTTGGGCTGTTTGTTTTACCTATGGCACATGGTTTGG 1213

Qy       1887 GGTAAAAGGTCTGATTGCTGCTGGAAGGAGTTTCAGTAATTGCTCAAGCATCCGTAAAGC 1946
Db        1214 TGTAAAGGGATTAGTTGCTGCTGGTAGGACATTCAAAAACAGTCCTGCAATCAGAAAGGC 1273

Qy       1947 TTGTGAATTTCTGCTGTCCAAGCAGCTTCCTTCTGGTGGCTGGGGAGAGAGTTATCTGTC 2006
Db        1274 ATGTGACTTTTTGTTGTCAAAAGAGCTTCCTTCTGGAGGCTGGGGAGAAAGCTATTTGTC 1333

Qy       2007 CTGTCAAAACAAGGTGTATTCAAATCTGGAAGGCAACAGGTCTCATGTGGTCAACACTGG 2066
Db        1334 ATCCCAAGATCAGGTTTATACCAATCTCGAAGGGAAGCGACCTCATGCGGTGAACACTGG 1393

Qy       2067 GTGGGCTATGTTGGCTCTCATTGATGCTGGACAGGCTAAGAGAGATTCGCAACCACTGCA 2126
Db        1394 TTGGGCCATGCTAGCCCTAATCGATGCAGGGCAGGCTGAGAGAGATCCAATTCCTTTGCA 1453

Qy       2127 CCGGGCAGCTGCATACTTGATAAATTCCCAATTGGAGGACGGTGACTTTCCGCAGCAGGA 2186
Db        1454 TCGAGCAGCGAAGGTTTTGATCAACTTACAATCGGAAGATGGTGAATTTCCCCAGCAAGA 1513

Qy       2187 AATAATGGGAGTCTTCAACAAGAATTGCATGATCACATACGCCGCATACAGAAACATATT 2246
Db        1514 GATCATTGGAGTCTTCAACAAAAACTGCATGATCAGCTACTCCGAGTATAGAAACATCTT 1573

Qy       2247 CCCCATTTGGGCGTTGGGAGAATACCAATCCCAAGTATTG 2286
Db        1574 CCCTATTTGGGCCCTTGGGGAGTACCGTCGCCGCGTCTTG 1613
```

```
RESULT 9
ADQ95185
ID   ADQ95185 standard; cDNA; 1882 BP.
XX
AC   ADQ95185;
XX
DT   23-SEP-2004  (first entry)
XX
DE   Rice cycloartenol synthase cDNA.
XX
KW   rice; cycloartenol synthase; squalene metabolism; gene; ss.
XX
OS   Oryza sativa.
XX
FH   Key             Location/Qualifiers
FT   CDS             3. .1628
FT                   /*tag=  a
FT                   /product= "Rice cycloartenol synthase"
FT                   /partial
FT                   /note= "No start codon given"
XX
PN   US2004121439-A1.
XX
PD   24-JUN-2004.
XX
PF   27-JUN-2003; 2003US-00607726.
XX
PR   23-OCT-1998;   98US-0105405P.
PR   15-OCT-1999;   99US-00419679.
XX
PA   (FALC/) FALCO S C.
PA   (FAMO/) FAMODU O O.
PA   (HITZ/) HITZ W D.
PA   (KINN/) KINNEY A J.
PA   (RAFA/) RAFALSKI J A.
PA   (MCGO/) MCGONIGLE B.
PA   (LOHM/) LOHMAN K N.
XX
PI   Falco SC,  Famodu OO,  Hitz WD,  Kinney AJ,  Rafalski JA;
PI   Mcgonigle B,  Lohman KN;
XX
DR   WPI; 2004-479684/45.
DR   P-PSDB; ADQ95186.
XX
PT   New composition comprising enzymes involved in squalene metabolism,
PT   useful in altering squalene metabolism in plants and seeds.
XX
PS   Claim 2; SEQ ID NO 3; 33pp; English.
XX
CC   The invention describes a composition comprising an isolated
CC   polynucleotide comprising a nucleotide sequence encoding a first
CC   polypeptide of at least 200 amino acids. A composition comprising an
CC   isolated polynucleotide comprising a nucleotide sequence encoding a first
CC   polypeptide of at least 200 amino acids that has at least 90% identity
CC   based on the Clustal method of alignment when compared to a polypeptide
CC   comprising a sequence of 701, 541, 757, 345, 386, 330, 523, or 467 amino
CC   acids (EVEN SEQ ID NOS: 2-16), or an isolated polynucleotide comprising
CC   the complement of the nucleotide sequence, or comprising a polypeptide of
CC   at least 200 amino acids that has at least 90% identity based on the
CC   Clustal method of alignment when compared to a polypeptide of EVEN SEQ ID
CC   NOS: 2-16. Also described are: a chimeric gene comprising the isolated
CC   polynucleotide described above operably linked to suitable regulatory
CC   sequences; an isolated host cell comprising the chimeric gene of (1) or
CC   the isolated polynucleotide described above; a virus comprising the
CC   isolated polynucleotide described above; selecting an isolated
CC   polynucleotide that affects the level of expression of a polypeptide in a
CC   host or plant cell; obtaining a nucleic acid fragment encoding a
CC   polypeptide; and evaluating at least one compound for its ability to
CC   inhibit the activity of a polypeptide. The composition, methods,
CC   polynucleotides and polypeptides are useful in altering squalene
CC   metabolism in plants and seeds. This sequence encodes corn cycloartenol
CC   synthase, a member of the squalene metabolism pathway.
XX
```

```
Qy       1574 ATGTCATTCTCTCACTACAGAATGAAGATGGTGGTTTTGCAACATATGAGCTTAAACGAT 1633
             ||||| |  | || || |||| | || | |||||||| ||| |||  |||| |  | ||
Db       1589 ATGTCTTGTTGTCGCTTCAGAGTAAAAAAGGTGGTTTAGCAGCATGGGAGCCTGCAGGAG 1648

Qy       1634 CTTATAATTGGTTGGAGATAATCAATCCTGCTGAAACTTTTGGTGACATCGTTATTGATT 1693
             || |  | ||||| ||  || |||||||  | |||  |||||  ||||| ||  ||||
Db       1649 CTCAAGAGTGGTTAGAATTACTCAATCCCACAGAATTTTTTGCGGACATTGTAGTTGAAC 1708

Qy       1694 ATCCTTATGTGGAATGTACATCAGCAGCGATTCAAGCTTTGGCATCATTTAGGAAATTAT 1753
             ||   ||||| || || ||   | | || || |||||||| |  |  || | |||  |||
Db       1709 ATGAATATGTTGAGTGCACTGGATCTGCAATCCAAGCTTTAGTTTTGTTCAAGAAGCTAT 1768

Qy       1754 ATCCTGGGCATCGCCGAGAAGAAATACAACATTGTATCGATAAAGCCACTACCTTCATTG 1813
             |||| || ||| |      |||| ||| || |||  ||    || ||    |  ||| |||
Db       1769 ATCCAGGACATAGGAAGAAAGAGATAGAAAATTTCATTACCAATGCAGTTCGATTCCTTG 1828

Qy       1814 AAAAAATACAAGCTTCAGATGGATCATGGTATGGTTCTTGGGGAGTTTGCTTCACTTACG 1873
             || | | |||| |  | ||||| ||||||||||  |||||||||||||||||||| |
Db       1829 AAGATACACAAACAGCTGATGGTTCATGGTATGGAAATTGGGGAGTTTGCTTCACTTATG 1888

Qy       1874 GTGCTTGGTTTGGGGTAAAAGGTCTGATTGCTGCTGGAAGGAGTTTCAGTAATTGCTCAA 1933
             |  |||||||||   |  ||||||   ||||||| | || || ||  |||||  |
Db       1889 GCTCTTGGTTTGCACTTGGAGGTCTAGCAGCTGCTGGTAAGACTTACACCAATTGTGCTG 1948

Qy       1934 GCATCCGTAAAGCTTGTGAATTTCTGCTGTCCAAGCAGCTTCCTTCTGGTGGCTGGGGAG 1993
              ||| || |||||   | ||||||| ||  | |  ||          ||||| |||||||
Db       1949 CCATTCGCAAAGCCGTTAAATTTCTACTTACAACACAAAGAGAGGACGGTGGATGGGGAG 2008

Qy       1994 AGAGTTATCTGTCCTGTCAAAACAAGGTGTATTCAAATCTGGAAGGCAACAGGTCTCATG 2053
             |||||||||| ||  | | ||| ||| | |||  |  ||| ||||| | | | ||  |||
Db       2009 AGAGTTATCTTTCAAGCCCAAAAAAGATATATGTACCTCTAGAAGGAAGCCGATCAAATG 2068

Qy       2054 TGGTCAACACTGGGTGGGCTATGTTGGCTCTCATTGATGCTGGACAGGCTAAGAGAGATT 2113
             | ||  | || |  |||||| |  |||  || ||| ||||||||||||  | |||||
Db       2069 TTGTACATACAGCATGGGCTCTTATGGGACTAATTCATGCTGGACAGGCGGATAGAGACC 2128

Qy       2114 CGCAACCACTGCACCGGGCAGCTGCATACTTGATAAATTCCCAATTGGAGGACGGTGACT 2173
             |    || || ||||| || ||    |  | || ||||| || ||||| || ||||| |
Db       2129 CCATGCCTCTTCACCGTGCTGCAAAGTTGCTCATTAATTCTCAGTTGGAAGAGGGTGATT 2188

Qy       2174 TTCCGCAGCAGGAAATAATGGGAGTCTTCAACAAGAATTGCATGATCACATACGCCGCAT 2233
               || || |||||||| | |||||| |||| || |||||||| |    ||  |    |
Db       2189 GGCCCCAACAGGAAATCACGGGAGTATTCATGAAAAATTGCATGTTGCATTATCCAATGT 2248

Qy       2234 ACAGAAACATATTCCCCATTTGGGCGTTGGGAGAATACCAATCCCAAGT 2282
             ||||| | || |  || || |||||  | |  ||||| | |    |  ||
Db       2249 ACAGAGATATTTATCCAATGTGGGCTCTAGCTGAATATCGAAGGCGGGT 2297
```

RESULT 8
ADE28074
ID ADE28074 standard; cDNA; 1882 BP.
XX
AC ADE28074;
XX
DT 29-JAN-2004 (first entry)
XX
DE Rice cycloartenol synthase cDNA.
XX
KW squalene metabolism; transgenic; plant; genetic; physical gene mapping;
KW breeding; rice; cycloartenol synthase; ss; gene.
XX
OS Oryza sativa.
XX
FH Key Location/Qualifiers
FT CDS 3. .1628
FT /*tag= a
FT /partial
FT /product= "Rice cycloartenol synthase protein"
FT /note= "No start codon"
XX
PN US6630617-B1.
XX
PD 07-OCT-2003.
XX
PF 15-OCT-1999; 99US-00419679.
XX
PR 23-OCT-1998; 98US-0105405P.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Famodu OO, Kinney AJ, Rafalski JA;
XX
DR WPI; 2003-810554/76.
DR P-PSDB; ADE28075.
XX
PT New isolated polynucleotide encoding a squalene metabolic enzyme, useful
PT as probes in gene mapping, in producing transgenic plants or in plant
PT breeding to develop lines having desired phenotypes.
XX
PS Disclosure; SEQ ID NO 3; 31pp; English.
XX
CC The invention relates to a novel isolated polynucleotide encoding a
CC squalene metabolic enzyme. The polynucleotide of the invention may be
CC used to create transgenic plants in which the polypeptides are present at
CC higher or lower levels than normal or are found within cell types or
CC developmental stages in which they are not normally found. Furthermore,
CC the polynucleotide may be used as a probe for genetically and physically
CC mapping genes and as a marker for traits linked to those genes, as well
CC as in plant breeding in order to develop lines with desired phenotypes.
CC The current sequence is that of the rice cycloartenol synthase cDNA of
CC the invention.
XX
SQ Sequence 1882 BP; 565 A; 363 C; 441 G; 513 T; 0 U; 0 Other;

Query Match 36.9%; Score 944; DB 10; Length 1882;
Best Local Similarity 74.4%; Pred. No. 9e-247;
Matches 1190; Conservative 0; Mismatches 410; Indels 0; Gaps 0;

```
Qy        687 GCTTTCAGTACTTGGAGTGTATGAATGGTCTGGAAATAATCCCCTGCCCCCTGAGATATG 746
              |||||| || ||||| || | ||| |||||||| || || ||  |||| || || |||||
Db         14 GCTTTCGGTGCTTGGTGTATTTGACTGGTCTGGCAACAACCCAGTGCCACCAGAAATATG 73

Qy        747 GCTCCTTCCATACATGCTTCCATTTCATCCAGGAAGGATGTGGTGTCACTGCCGGATGGT 806
              | |  | |||||  | || ||  |||||||||  | ||||||||||| ||||||||||
Db         74 GTTGTTGCCATATTTCCTGCCGATTCATCCAGGGCGAATGTGGTGTCATTGCCGGATGGT 133

Qy        807 CTATTTGCCGATGTCCTACTTATATGGCAAGAGGTTTGTTGGTCCAATCTCACCAACAGT 866
              |||||||| ||||  ||| | ||||| ||||||||||| || ||| |  ||||||   |
Db        134 TTATTTGCCTATGTGTTACATTTATGGAAAGAGGTTTGTGGGCCCAGTTACACCAATTAT 193

Qy        867 ATTATCTTTGAGAAAAGAGCTTTATACAGTACCATACCATGATATAGATTGGGATCAGGC 926
              |||    || ||||| || || ||   |||||| ||| ||||  | ||||||||  ||||
Db        194 ATTGGAATTAAGAAAGGAACTCTACGAAGTACCCTACAATGAAGTTGATTGGGACAAGGC 253

Qy        927 TCGCAATTTGTGTGCAAAGGAAGATTTGTACTATCCTCACCCACTTGTACAGGATATTCT 986
              ||||||| | ||||| |||||||| |||||||||| || ||| | || |||||| |  |
Db        254 TCGCAATCTATGTGCTAAGGAAGATCTGTACTATCCACATCCATTCGTGCAGGATGTATT 313

Qy        987 TTGGGCATCTCTACACAAGTTCCTTGAGCCTATTCTGATGCATTGGCCTGGAAAAAGATT 1046
              |||||  |||| |||| ||  |||| || |  | || ||| |||||||||| || | |||
Db        314 ATGGGCCACTCTCCACAAATTTGTTGAACCAGCTATGTTGCGTTGGCCTGGGAACAAATT 373

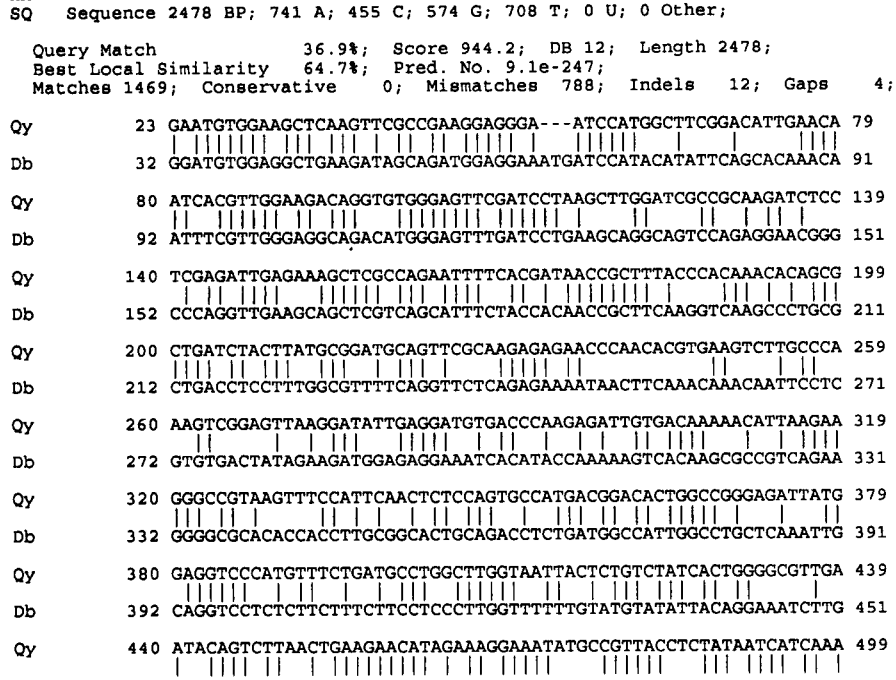
Qy       1047 GAGGGAAAAGGCTATTATTTCTGCATTGGAGCATATACATTACGAAGATGAGAATACTCG 1106
              |||||| || ||| |      |||   || ||||||| ||||| |||||||||| || ||
Db        374 GAGGGAGAAAGCTTTGGACACTGTCATGCAGCATATTCATTATGAAGATGAGAACACCCG 433

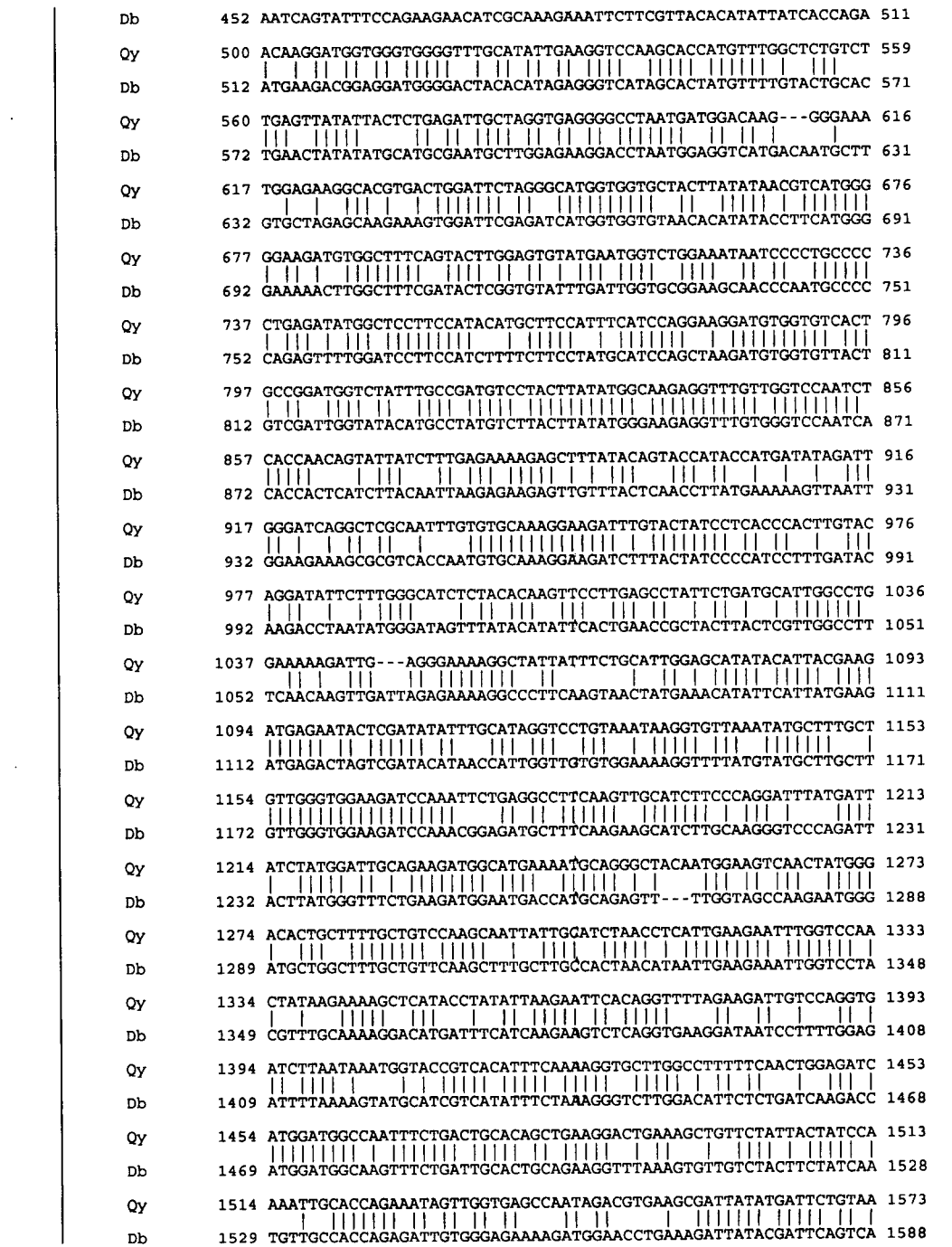
Qy       1107 ATATATTTGCATAGGTCCTGTAAATAAGGTGTTAAATATGCTTTGCTGTTGGGTGGAAGA 1166
              |||||||||||| ||||| ||||| ||||||||||||   || ||| | |||||
Db        434 ATATATTTGCATTGGTCCAGTAAACAAGGTATTAAATATGCTTGCTTGCTGGATTGAAGA 493

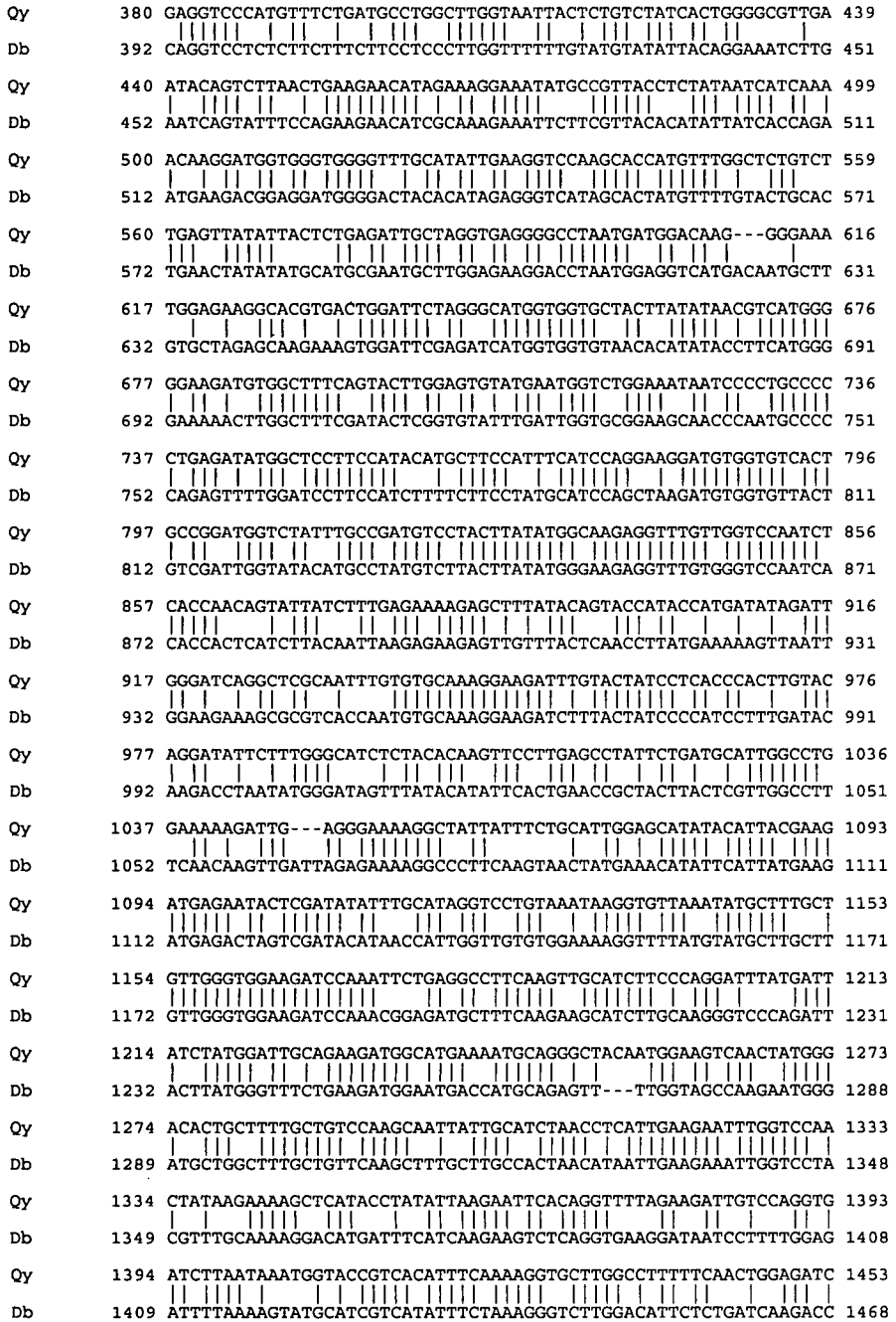
Qy       1167 TCCAAATTCTGAGGCCTTCAAGTTGCATCTTCCCAGGATTTATGATTATCTATGGATTGC 1226
              |||||| || ||||| ||||| | ||  |||| ||  |  | ||||| ||||||||||
Db        494 TCCAAACTCAGAGGCATTCAAACTCCACATTCCAAGAGTCCACGATTACCTATGGATTGC 553
```

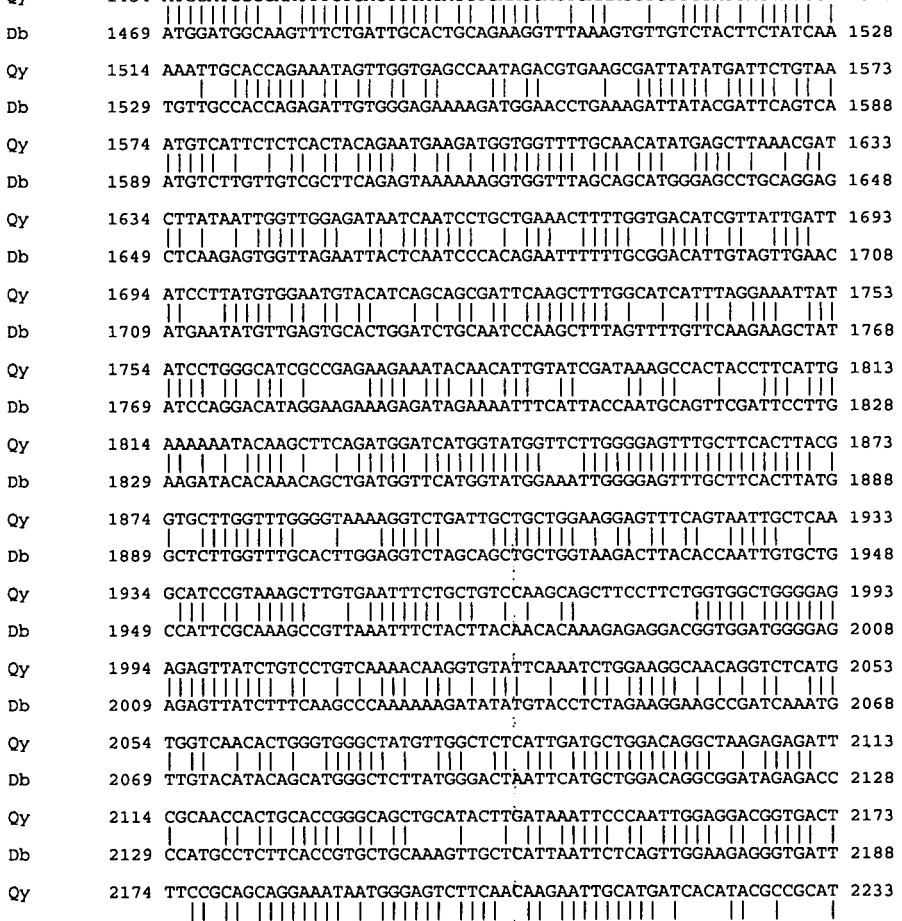
XX
PD 20-NOV-2003.
XX
PF 05-MAY-2003; 2003WO-US014218.
XX
PR 09-MAY-2002; 2002US-0379361P.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Mcgonigle B, Maxwell CA, Hession AO;
XX
DR WPI; 2004-012098/01.
XX
PT New transgenic plants comprising recombinant DNA molecules encoding
PT oxidosqualene cyclase to lower triterpene level, used as feeds, foods,
PT beverages and industrial products like agricultural adjuvants or concrete
PT supplies.
XX
PS Example 1; SEQ ID NO 4; 39pp; English.
XX
CC The invention relates to a plant comprising at least one recombinant DNA
CC molecule comprising a promoter operably linked to at least a portion of
CC at least one oxidosqualene cyclase gene, the molecule sufficient to
CC suppress the production of a triterpene or any of its progeny, where the
CC progeny comprise the molecule. The oxidosqualene cyclase gene catalyzes
CC the cyclization of 2,3-oxidosqualene to form a triterpene, e.g. beta-
CC amyrin, lanosterol, lupeol, cycloartenol, alpha-amyrin, isomultiflorenol
CC or any combination. The plants are useful in generating products which
CC may be used as feeds, foods, beverages and industrial products like
CC agricultural adjuvants, concrete supplies, dielectric fluids, dust
CC suppressants, fuel additives, hydraulic fluids, industrial cleaners,
CC industrial lubricants, metalworking fluids, odor reduction, paint
CC strippers, printing inks, printing supplies or saw guide oils. Saponins
CC may play a defense role against pathogens in plant tissues. The present
CC sequence represents a cDNA insert from clone src3c.pk024.m11 encoding a
CC G. max beta-amyrin synthase.
XX
SQ Sequence 2478 BP; 741 A; 455 C; 574 G; 708 T; 0 U; 0 Other;

Query Match 36.9%; Score 944.2; DB 12; Length 2478;
Best Local Similarity 64.7%; Pred. No. 9.1e-247;
Matches 1469; Conservative 0; Mismatches 788; Indels 12; Gaps 4;









RESULT 7
ADF89882
ID ADF89882 standard; cDNA; 2478 BP.
XX
AC ADF89882;
XX
DT 26-FEB-2004 (first entry)
XX
DE G. max beta-amyrin synthase cDNA fragment (clone src3c.pk024.m11).
XX
KW Oxidosqualene cyclase; triterpene; plant protectant; industrial product;
KW plant; soybean; gene; ss; beta-amyrin synthase.
XX
OS Glycine max.
XX
PN WO2003095615-A2.

```
Qy       1564 GATTCTGTAAATGTCATTCTCTCACTACAGAATGAAGATGGTGGTTTTGCAACATATGAG 1623
Db       1383 GATGCTGTCAGTTGTCTGATGTCTTATATGAATGATAATGGCGGTTTCGCGACATATGAA 1442

Qy       1624 CTTAAACGATCTTATAATTGGTTGGAGATAATCAATCCTGCTGAAACTTTTGGTGACATC 1683
Db       1443 CTCACAAGATCTTATCCCTGGTTGGAGCTGATCAATCCCGCTGAGACCTTTGGGGATATT 1502

Qy       1684 GTTATTGATTATCCTTATGTGGAATGTACATCAGCAGCGATTCAAGCTTTGGCATCATTT 1743
Db       1503 GTGATTGATTACCCGTATGTTGAATGTACATCAGCAGCAATTCAGGCCCTGACATCATTC 1562

Qy       1744 AGGAAATTATATCCTGGGCATCGCCGAGAAGAAATACAACATTGTATCGATAAAGCCACT 1803
Db       1563 AAAAAACTATACCCTGGGCACCGCAGGAAAGAGGTGGATAACTGTATCAGCAAAGCTTCC 1622

Qy       1804 ACCTTCATTGAAAAAATACAAGCTTCAGATGGATCATGGTATGGTTCTTGGGGAGTTTGC 1863
Db       1623 AATTTCATCGAGAGTATTCAGAAAAGCGATGGTTCATGGTATGGCTCTTGGGCCGTCTGT 1682

Qy       1864 TTCACTTACGGTGCTTGGTTTGGGGTAAAAGGTCTGATTGCTGCTGGAAGGAGTTTCAGT 1923
Db       1683 TTCACATACGGCACTTGGTTTGGTGTGAAGGGACTAATTGCTGCTGGTAGAACATTTGAG 1742

Qy       1924 AATTGCTCAAGCATCCGTAAAGCTTGTGAATTTCTGCTGTCCAAGCAGCTTCCTTCTGGT 1983
Db       1743 AACAGTCCTGCAATTAGAAAGGCATGCGACTTTCTGTTGTCAAAAGAACTTCCTTCCGGT 1802

Qy       1984 GGCTGGGGAGAGAGTTATCTGTCCTGTCAAAACAAGGTGTATTCAAATCTGGAAGGCAAC 2043
Db       1803 GGTTGGGGAGAAAGCTATTTGTCATCTCAAGACCAGGTTTACACCAATCTCAAAGGCAAC 1862

Qy       2044 AGGTCTCATGTGGTCAACACTGGGTGGGCTATGTTGGCTCTCATTGATGCTGGACAGGCT 2103
Db       1863 CGGCCTCATGCGGTGAACACTAGTTGGGCCATGCTGGCGCTGATTGATGCGGGCCAGGCC 1922

Qy       2104 AAGAGAGATTCGCAACCACTGCACCGGGCAGCTGCATACTTGATAAATTCCCAATTGGAG 2163
Db       1923 GAGAGAGATCCAACGCCTCTACACCGAGCAGCAAAGGTTTTGATCAACTTACAATCAGAG 1982

Qy       2164 GACGGTGACTTTCCGCAGCAGGAAATAATGGGAGTCTTCAACAAGAATTGCATGATCACA 2223
Db       1983 GACGGAGAATTTCCTCAGCAAGAGATCATAGGAGTGTTCAACAAGAACTGCATGATAAGC 2042

Qy       2224 TACGCCGCATACAGAAACATATTCCCCATTTGGGCGTTGGGAGAATACCAATCCCAAGTA 2283
Db       2043 TACTCCCAGTACAGGAACATCTTCCCGATTTGGGCTCTGGGTGAGTACCGGTGTCGAGTC 2102

Qy       2284 TTGCAATCTCGTTAATCGAGCCTTAGTTAGGGTG 2317
Db       2103 TTGGGGGCTGGCAAGCCTTGGCGGTGAACGGGAG 2136
```

```
RESULT 6
AAD16894
ID   AAD16894 standard; cDNA; 2478 BP.
XX
AC   AAD16894;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Soybean beta-amyrin synthase protein cDNA from clone src3c.pk024.m11.
XX
KW   Soybean; oxidosqualene cyclase; cytostatic; antilipaemic; colon cancer;
KW   transgenic plant; insect tolerance; soy flour; bean; nut; tofu; natto;
KW   miso; tempeh; beta-amyrin synthase; ss.
XX
OS   Glycine max.
XX
FH   Key             Location/Qualifiers
FT   CDS             34. .2322
FT                   /*tag=  a
FT                   /product= "Soybean beta-amyrin synthase protein"
XX
PN   WO200166773-A2.
XX
PD   13-SEP-2001.
XX
PF   09-MAR-2001; 2001WO-US007611.
XX
PR   09-MAR-2000; 2000US-0188054P.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Bryan GT,  Mcgonigle B,  Maxwell CA,  Potter SM,  Hwang D;
XX
DR   WPI; 2001-565594/63.
DR   P-PSDB; AAE09795.
XX
PT   New nucleic acid fragment encoding oxidosqualene cyclase, useful as DNA
PT   hybridization probe, polymerase chain reaction primer and for creating
PT   transgenic plants with cells having higher or lower of oxidosqualene
PT   cyclase.
XX
PS   Claim 2; Page 44-45; 59pp; English.
XX
CC   The invention relates to oxidosqualene cyclase proteins and their
CC   corresponding cDNA molecules. Oxidosqualene cyclase genes are used as
CC   hybridisation probes to screen libraries from any desired plant.
CC   Oxidosqualene cyclase DNA are used to create transgenic plants with cells
CC   having higher or lower level of oxidosqualene cyclase. These changes in
CC   the plant seed are useful to improve the seed nutritional value, and in
CC   the plant leaf may aid in insect tolerance. Oxidosqualene cyclase are
CC   produced in heterologous host cells and are used to prepare antibodies.
CC   The antibodies are useful for detecting oxidosqualene cyclase in situ in
CC   cells or in vitro in cell extracts. The invention is useful for altering
CC   level of expression of oxidosqualene cyclase (beta-amyrin synthase)
CC   protein in a plant cell. Soybean grain produced from a transgenic plant
CC   is useful for preparing food products such as soy protein product,
CC   soybean meal, soy flour, soy protein concentrate, soy milk, a dietary
CC   supplement, beans, nuts, tofu, natto, miso and tempeh. The formed soy
CC   protein product having increased saponin levels is useful for lowering
CC   cholesterol in a person suspected of having increased cholesterol, and
CC   for decreasing risk of colon cancer in a person suspected of having risk
CC   of cancer. The present cDNA sequence encodes soybean beta-amyrin synthase
CC   protein
XX
SQ   Sequence 2478 BP; 741 A; 455 C; 574 G; 708 T; 0 U; 0 Other;
```

```
  Query Match             36.9%;  Score 944.2;  DB 4;  Length 2478;
  Best Local Similarity   64.7%;  Pred. No. 9.1e-247;
  Matches 1469;  Conservative    0;  Mismatches  788;  Indels   12;  Gaps    4;

Qy         23 GAATGTGGAAGCTCAAGTTCGCCGAAGGAGGGA---ATCCATGGCTTCGGACATTGAACA 79
Db         32 GGATGTGGAGGCTGAAGATAGCAGATGGAGGAAATGATCCATACATATTCAGCACAAACA 91

Qy         80 ATCACGTTGGAAGACAGGTGTGGGAGTTCGATCCTAAGCTTGGATCGCCGCAAGATCTCC 139
Db         92 ATTTCGTTGGGAGGCAGACATGGGAGTTTGATCCTGAAGCAGGCAGTCCAGAGGAACGGG 151

Qy        140 TCGAGATTGAGAAAGCTCGCCAGAATTTTCACGATAACCGCTTTACCCACAAACACAGCG 199
Db        152 CCCAGGTTGAAGCAGCTCGTCAGCATTTCTACCACAACCGCTTCAAGGTCAAGCCCTGCG 211

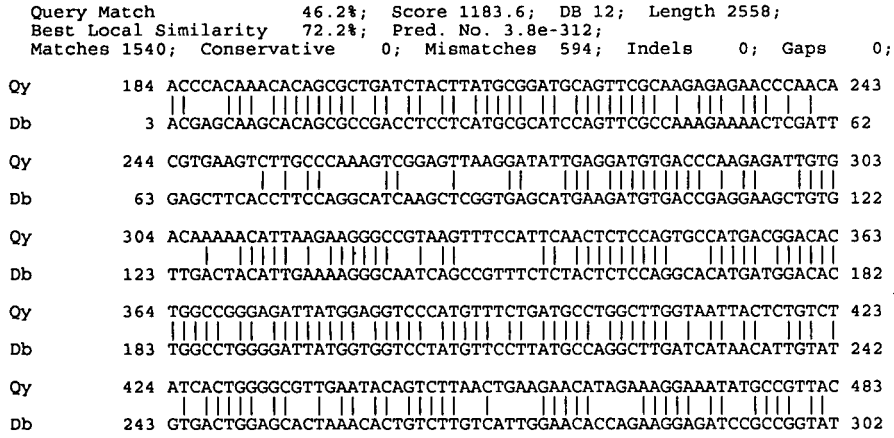
Qy        200 CTGATCTACTTATGCGGATGCAGTTCGCAAGAGAGAACCCAACACGTGAAGTCTTGCCCA 259
Db        212 CTGACCTCCTTTGGCGTTTTCAGGTTCTCAGAGAAAATAACTTCAAACAAACAATTCCTC 271

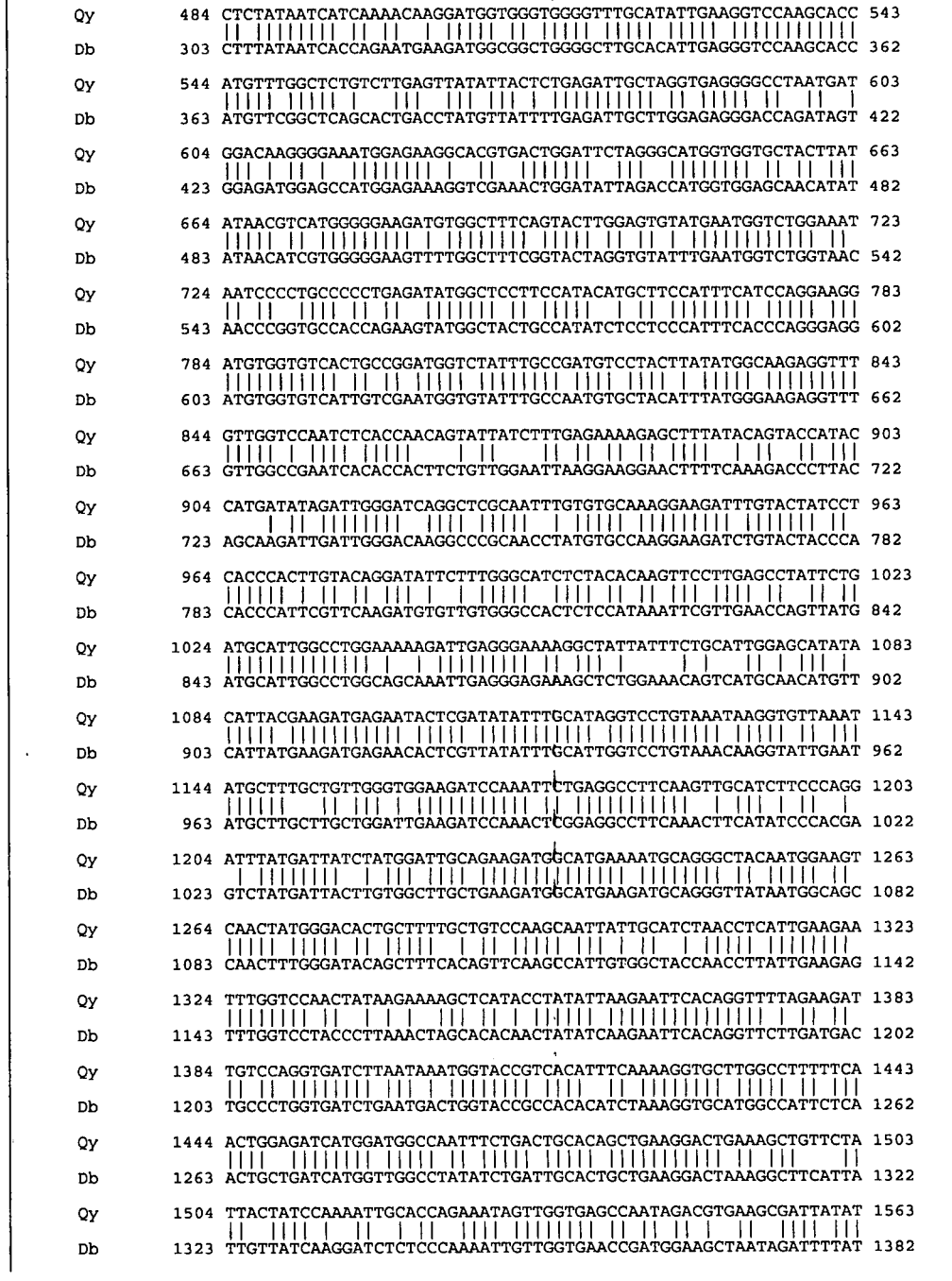
Qy        260 AAGTCGGAGTTAAGGATATTGAGGATGTGACCCAAGAGATTGTGACAAAAACATTAAGAA 319
Db        272 GTGTGACTATAGAAGATGGAGAGGAAATCACATACCAAAAAGTCACAAGCGCCGTCAGAA 331

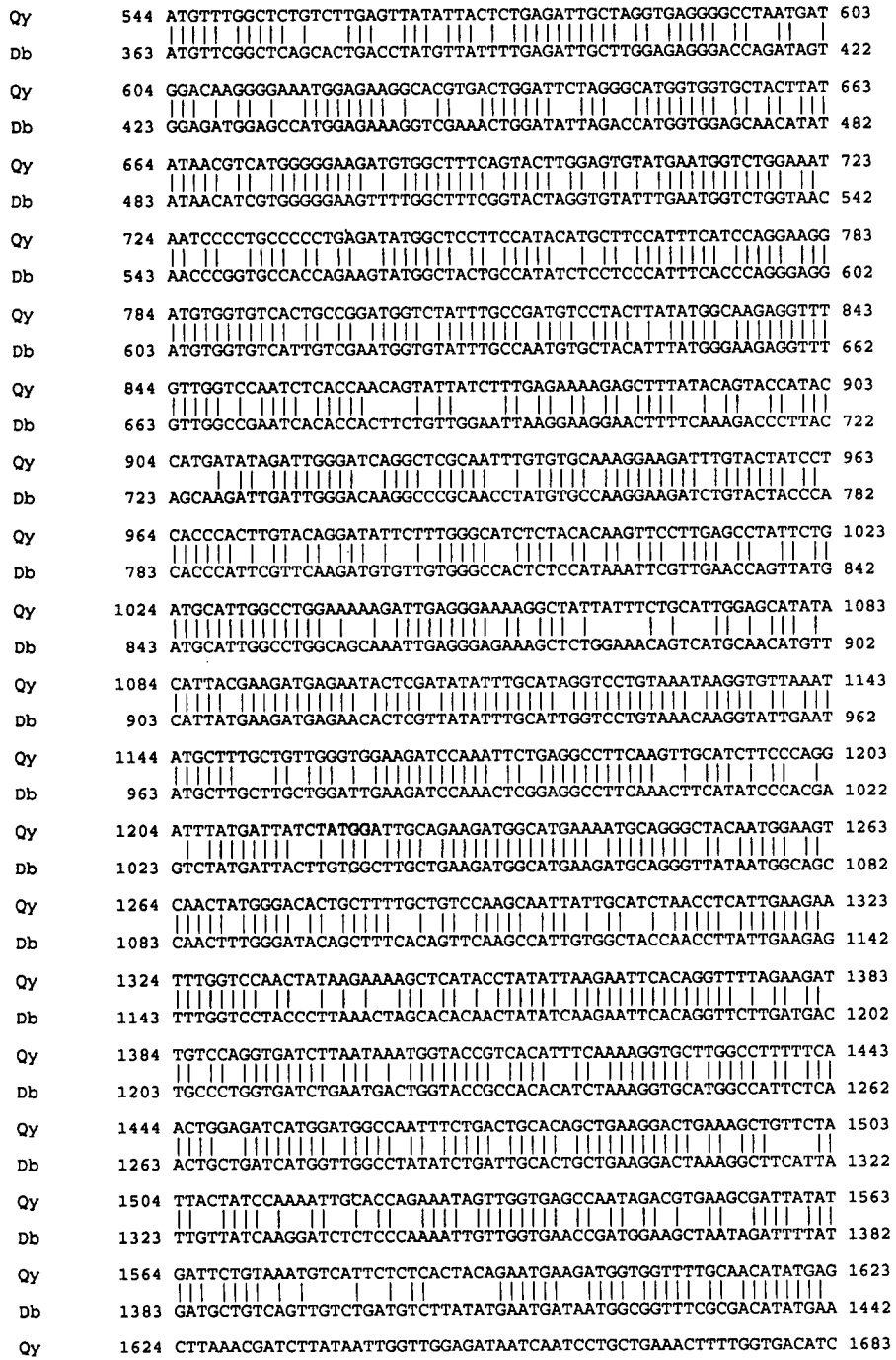
Qy        320 GGGCCGTAAGTTTCCATTCAACTCTCCAGTGCCATGACGGACACTGGCCGGGAGATTATG 379
Db        332 GGGGCGCACACCACCTTGCGGCACTGCAGACCTCTGATGGCCATTGGCCTGCTCAAATTG 391
```

PF 27-JUN-2003; 2003US-00607726.
XX
PR 23-OCT-1998; 98US-0105405P.
PR 15-OCT-1999; 99US-00419679.
XX
PA (FALC/) FALCO S C.
PA (FAMO/) FAMODU O O.
PA (HITZ/) HITZ W D.
PA (KINN/) KINNEY A J.
PA (RAFA/) RAFALSKI J A.
PA (MCGO/) MCGONIGLE B.
PA (LOHM/) LOHMAN K N.
XX
PI Falco SC, Famodu OO, Hitz WD, Kinney AJ, Rafalski JA;
PI Mcgonigle B, Lohman KN;
XX
DR WPI; 2004-479684/45.
DR P-PSDB; ADQ95184.
XX
PT New composition comprising enzymes involved in squalene metabolism,
PT useful in altering squalene metabolism in plants and seeds.
XX
PS Claim 2; SEQ ID NO 1; 33pp; English.
XX
CC The invention describes a composition comprising an isolated
CC polynucleotide comprising a nucleotide sequence encoding a first
CC polypeptide of at least 200 amino acids. A composition comprising an
CC isolated polynucleotide comprising a nucleotide sequence encoding a first
CC polypeptide of at least 200 amino acids that has at least 90% identity
CC based on the Clustal method of alignment when compared to a polypeptide
CC comprising a sequence of 701, 541, 757, 345, 386, 330, 523, or 467 amino
CC acids (EVEN SEQ ID NOS: 2-16), or an isolated polynucleotide comprising
CC the complement of the nucleotide sequence, or comprising a polypeptide of
CC at least 200 amino acids that has at least 90% identity based on the
CC Clustal method of alignment when compared to a polypeptide of EVEN SEQ ID
CC NOS: 2-16. Also described are: a chimeric gene comprising the isolated
CC polynucleotide described above operably linked to suitable regulatory
CC sequences; an isolated host cell comprising the chimeric gene of (1) or
CC the isolated polynucleotide described above; a virus comprising the
CC isolated polynucleotide described above; selecting an isolated
CC polynucleotide that affects the level of expression of a polypeptide in a
CC host or plant cell; obtaining a nucleic acid fragment encoding a
CC polypeptide; and evaluating at least one compound for its ability to
CC inhibit the activity of a polypeptide. The composition, methods,
CC polynucleotides and polypeptides are useful in altering squalene
CC metabolism in plants and seeds. This sequence encodes corn cycloartenol
CC synthase, a member of the squalene metabolism pathway.
XX
SQ Sequence 2558 BP; 706 A; 532 C; 636 G; 684 T; 0 U; 0 Other;

Query Match 46.2%; Score 1183.6; DB 12; Length 2558;
Best Local Similarity 72.2%; Pred. No. 3.8e-312;
Matches 1540; Conservative 0; Mismatches 594; Indels 0; Gaps 0;








```
RESULT 5
ADQ95183
ID   ADQ95183 standard; cDNA; 2558 BP.
XX
AC   ADQ95183;
XX
DT   23-SEP-2004  (first entry)
XX
DE   Corn cycloartenol synthase cDNA.
XX
KW   corn; cycloartenol synthase; squalene metabolism; gene; ss.
XX
OS   Zea mays.
XX
FH   Key             Location/Qualifiers
FT   CDS             3. .2105
FT                   /*tag=  a
FT                   /product= "Corn cycloartenol synthase"
FT                   /partial
FT                   /note= "No start or stop codon given"
XX
PN   US2004121439-A1.
XX
PD   24-JUN-2004.
XX
```

```
Qy       1405 TGGTACCGTCACATTTCAAAAGGTGCTTGGCCTTTTTCAACTGGAGATCATGGATGGCCA 1464
             ||||| || |||||||| ||||| |||||||||||| ||||||| |||||| || |||||
Db       1381 TGGTATCGCCACATTTCTAAAGGGGCTTGGCCTTTCTCAACTGCAGATCACGGTTGGCCC 1440

Qy       1465 ATTTCTGACTGCACAGCTGAAGGACTGAAAGCTGTTCTATTACTATCCAAAATTGCACCA 1524
             || |||||||||||  || ||||||||||||||| ||| || |||||||| || |
Db       1441 ATCTCTGACTGCACCGCAGAAGGACTGAAAGCTGCTCTTTTGCTATCCAAAGTTCCCAAG 1500

Qy       1525 GAAATAGTTGGTGAGCCAATAGACGTGAAGCGATTATATGATTCTGTAAATGTCATTCTC 1584
             |  || |||||||| |||||||| |  || || ||||||||  |||| ||||| ||  |
Db       1501 GCGATTGTTGGTGAACCAATAGATGCAAAACGGTTATATGAAGCTGTTAATGTTATCATT 1560

Qy       1585 TCACTACAGAATGAAGATGGTGGTTTTGCAACATATGAGCTTAAACGATCTTATAATTGG 1644
             ||  ||||||||| |||||| ||   | ||||||||||||||  |    | || ||   ||||
Db       1561 TCTTTACAGAATGCAGATGGAGGCCTCGCAACATATGAGCTCACCAGGTCATACCCTTGG 1620

Qy       1645 TTGGAGATAATCAATCCTGCTGAAACTTTTGGTGACATCGTTATTGATTATCCTTATGTG 1704
             || ||| ||||||| || || ||||| ||||| || || ||||||||||||||||| |||
Db       1621 TTAGAGCTAATCAACCCAGCAGAAACCTTTGGCGATATTGTTATTGATTATCCTTACGTG 1680

Qy       1705 GAATGTACATCAGCAGCGATTCAAGCTTTGGCATCATTTAGGAAATTATATCCTGGGCAT 1764
             |||||||||||||| || || |||||||||  ||||||| | ||  | |||||||| |||
Db       1681 GAATGTACATCAGCTGCTATCCAAGCTTTGATATCATTTCGAAAGCTGTATCCTGGTCAT 1740

Qy       1765 CGCCGAGAAGAAATACAACATTGTATCGATAAAGCCACTACCTTCATTGAAAAAATACAA 1824
             ||      | ||| || |  | || || || || ||    || |||||||||    || |||
Db       1741 CGAAAGAAGGAAGTAGATGAGTGCATTGAGAAGGCGGTTAAGTTCATTGAATCCATTCAA 1800

Qy       1825 GCTTCAGATGGATCATGGTATGGTTCTTGGGGAGTTTGCTTCACTTACGGTGCTTGGTTT 1884
             ||  ||||||| |||||||||| || ||||  |||||||||| || ||| | ||||||
Db       1801 GCAGCAGATGGCTCATGGTATGGATCATGGGCTGTTTGCTTCACGTATGGTACGTGGTTT 1860

Qy       1885 GGGGTAAAAGGTCTGATTGCTGCTGGAAGGAGTTTCAGTAATTGCTCAAGCATCCGTAAA 1944
             || || ||||| ||| | |||| |||||  |  || |  || |   ||     |   ||||
Db       1861 GGAGTGAAAGGGCTGGTAGCTGTTGGAAAAACATTGAAAAACTCTCCACATGTTGCTAAA 1920

Qy       1945 GCTTGTGAATTTCTGCTGTCCAAGCAGCTTCCTTCTGGTGGCTGGGGAGAGAGTTATCTG 2004
             |||||||||||||  |||| || || |  ||||| || |||||||||| || |||||
Db       1921 GCTTGTGAATTTCTATTGTCGAAACAACAACCTTCGGGCGGCTGGGGAGAAAGCTATCTT 1980

Qy       2005 TCCTGTCAAAACAAGGTGTATTCAAATCTGGAAGGCAACAGGTCTCATGTGGTCAACACT 2064
             || |||||| ||||||| |||||||| || || |||||||| ||||| || || || ||
Db       1981 TCATGTCAAGACAAGGTCTATTCAAACCTTGATGGCAACAGATCTCACGTCGTGAATACA 2040

Qy       2065 GGGTGGGCTATGTTGGCTCTCATTGATGCTGGACAGGCTAAGAGAGATTCGCAACCACTG 2124
             |  |||||||| | || ||||||| |||||| || ||| ||  |||   | |||||||
Db       2041 GCATGGGCTATGCTCGCACTCATTGGTGCTGGGCAAGCTGAGGTAGACCGGAAACCACTA 2100

Qy       2125 CACCGGGCAGCTGCATACTTGATAAATTCCCAATTGGAGGACGGTGACTTTCCGCAGCAG 2184
             |||||||| ||   ||||||||| ||| | ||| ||||| | ||||| ||||| || |||
Db       2101 CACCGGGCTGCAAGATACTTGATTAATGCTCAAATGGAGAATGGTGATTTTCCACAACAG 2160

Qy       2185 GAAATAATGGGAGTCTTCAACAAGAATTGCATGATCACATACGCCGCATACAGAAACATA 2244
             |||||||||||||||||||| | ||| |||||||| ||||| ||||| ||  |||||||
Db       2161 GAAATAATGGGAGTCTTCAATAGGAACTGCATGATAACATATGCCGCGTATCGAAACATT 2220

Qy       2245 TTCCCCATTTGGGCGTTGGGAGAATACCAATCCCAAGTATT 2285
             || || || ||||| ||||| || ||||  |  || |||||
Db       2221 TTTCCGATATGGGCTTTGGGGGAGTACCGTTGTCAGGTATT 2261
```

RESULT 4
ADE28072
ID ADE28072 standard; cDNA; 2558 BP.
XX
AC ADE28072;
XX
DT 29-JAN-2004 (first entry)
XX
DE Corn cycloartenol synthase cDNA.
XX
KW squalene metabolism; transgenic; plant; genetic; physical gene mapping;
KW breeding; corn; cycloartenol synthase; ss; gene; maize.
XX
OS Zea mays.
XX
FH Key Location/Qualifiers
FT CDS 3. .2129
FT /*tag= a
FT /partial
FT /product= "Corn cycloartenol synthase protein"
FT /note= "No start codon"
XX
PN US6630617-B1.
XX
PD 07-OCT-2003.
XX
PF 15-OCT-1999; 99US-00419679.
XX
PR 23-OCT-1998; 98US-0105405P.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Famodu OO, Kinney AJ, Rafalski JA;
XX
DR WPI; 2003-810554/76.
DR P-PSDB; ADE28073.
XX
PT New isolated polynucleotide encoding a squalene metabolic enzyme, useful
PT as probes in gene mapping, in producing transgenic plants or in plant
PT breeding to develop lines having desired phenotypes.
XX
PS Disclosure; SEQ ID NO 1; 31pp; English.
XX
CC The invention relates to a novel isolated polynucleotide encoding a
CC squalene metabolic enzyme. The polynucleotide of the invention may be
CC used to create transgenic plants in which the polypeptides are present at
CC higher or lower levels than normal or are found within cell types or
CC developmental stages in which they are not normally found. Furthermore,
CC the polynucleotide may be used as a probe for genetically and physically
CC mapping genes and as a marker for traits linked to those genes, as well
CC as in plant breeding in order to develop lines with desired phenotypes.
CC The current sequence is that of the corn cycloartenol synthase cDNA of
CC the invention.
XX
SQ Sequence 2558 BP; 706 A; 532 C; 636 G; 684 T; 0 U; 0 Other;

Query Match 46.2%; Score 1183.6; DB 10; Length 2558;
Best Local Similarity 72.2%; Pred. No. 3.8e-312;
Matches 1540; Conservative 0; Mismatches 594; Indels 0; Gaps 0;

```
Qy        184 ACCCACAAACACAGCGCTGATCTACTTATGCGGATGCAGTTCGCAAGAGAGAACCCAACA 243
              ||   ||| |||||||| || || || ||||| || |||||||| | ||| ||| | |
Db          3 ACGAGCAAGCACAGCGCCGACCTCCTCATGCGCATCCAGTTCGCCAAAGAAAACTCGATT 62

Qy        244 CGTGAAGTCTTGCCCAAAGTCGGAGTTAAGGATATTGAGGATGTGACCCAAGAGATTGTG 303
                  | | ||     ||     |    ||   ||| ||||||||| | ||   ||||
Db         63 GAGCTTCACCTTCCAGGCATCAAGCTCGGTGAGCATGAAGATGTGACCGAGGAAGCTGTG 122

Qy        304 ACAAAAACATTAAGAAGGGCCGTAAGTTTCCATTCAACTCTCCAGTGCCATGACGGACAC 363
                |  ||||| | ||||||  | ||       || |||||||||   ||||| ||||||
Db        123 TTGACTACATTGAAAAGGGCAATCAGCCGTTTCTCTACTCTCCAGGCACATGATGGACAC 182

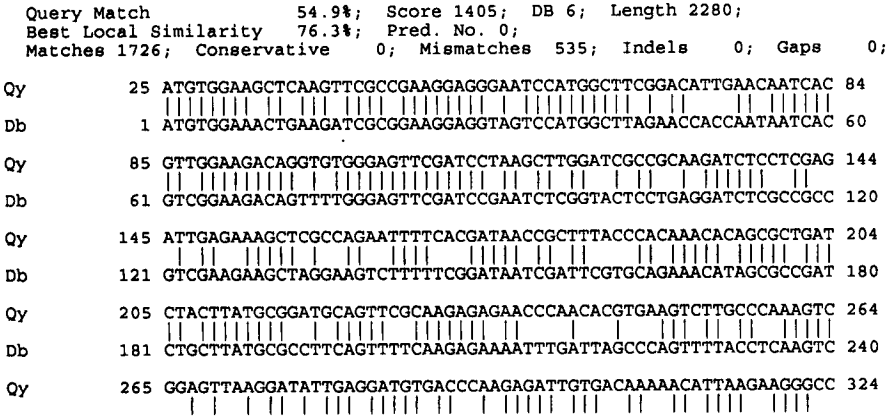
Qy        364 TGGCCGGGAGATTATGGAGGTCCCATGTTTCTGATGCCTGGCTTGGTAATTACTCTGTCT 423
              ||||| || |||||||| ||||| ||||| || ||||| |||||| | || ||  ||| |
Db        183 TGGCCTGGGGATTATGGTGGTCCTATGTTCCTTATGCCAGGCTTGATCATAACATTGTAT 242

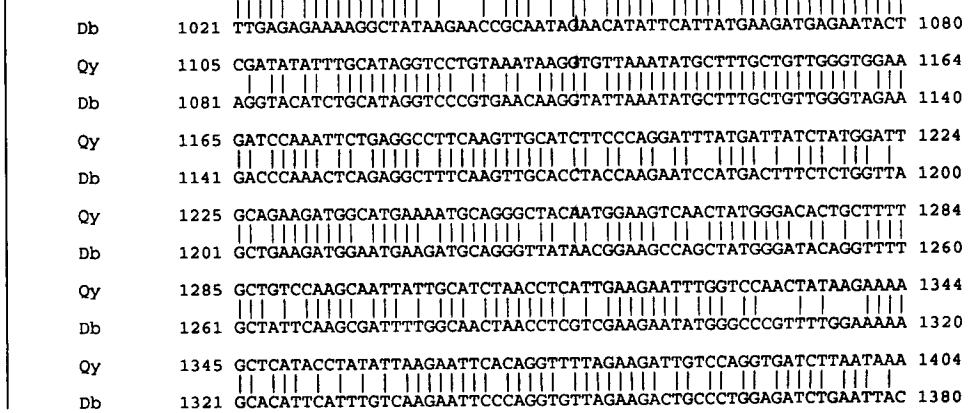
Qy        424 ATCACTGGGGCGTTGAATACAGTCTTAACTGAAGAACATAGAAAGGAAATATGCCGTTAC 483
              | ||||| ||  | || || |||||  |     |||||    ||||| ||  |||| ||
Db        243 GTGACTGGAGCACTAAACACTGTCTTGTCATTGGAACACCAGAAGGAGATCCGCCGGTAT 302

Qy        484 CTCTATAATCATCAAAACAAGGATGGTGGGTGGGGTTTGCATATTGAAGGTCCAAGCACC 543
              || |||||||| || ||  | ||||| || ||||| ||||| ||||||||||||
Db        303 CTTTATAATCACCAGAATGAAGATGGCGGCTGGGGCTTGCACATTGAGGGTCCAAGCACC 362
```

RESULT 3
ABZ13028
ID ABZ13028 standard; DNA; 2280 BP.
XX
AC ABZ13028;
XX
DT 21-JAN-2003 (first entry)
XX
DE Arabidopsis thaliana stress regulated gene SEQ ID NO 833.
XX
KW Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
XX
OS Arabidopsis thaliana.
XX
PN WO200216655-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US026685.
XX
PR 24-AUG-2000; 2000US-0227866P.
PR 26-JAN-2001; 2001US-0264647P.
PR 22-JUN-2001; 2001US-0300111P.
XX
PA (SCRI) SCRIPPS RES INST.
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Harper JF, Kreps J, Wang X, Zhu T;
XX
DR WPI; 2002-304127/34.
XX
PT Identifying a stress condition to which a plant cell has been exposed and
PT producing plants with increased tolerance to these abiotic stresses.
XX
PS Claim 144; SEQ ID NO 833; 577pp + Sequence Listing; English.
XX
CC The invention relates to identifying a stress condition to which a plant
CC cell has been exposed, comprising: (a) contacting nucleic acid
CC representative of expressed polynucleotides in the plant cell with an
CC array or probes representative of the plant cell genome; and (b)
CC detecting a profile of expressed polynucleotides in the plant cell
CC characteristic of a stress response. The method is useful in the
CC production of transgenic plants, cells and seeds and in producing plants
CC with increased tolerance to abiotic stress. The present sequence is that
CC of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used
CC in methods of the invention. Note: The sequence data for this patent is
CC not represented in the printed specification but is based on sequence
CC information supplied to Derwent by the European Patent Office
XX
SQ Sequence 2280 BP; 653 A; 449 C; 562 G; 616 T; 0 U; 0 Other;

Query Match 54.9%; Score 1405; DB 6; Length 2280;
Best Local Similarity 76.3%; Pred. No. 0;
Matches 1726; Conservative 0; Mismatches 535; Indels 0; Gaps 0;





```
Qy       421 TCTATCACTGGGGCGTTGAATACAGTCTTAACTGAAGAACATAGAAAGGAAATATGCCGT 480
Db       421 TCTATCACTGGGGCGTTGAATACAGTCTTAACTGAAGAACATAGAAAGGAAATATGCCGT 480

Qy       481 TACCTCTATAATCATCAAAACAAGGATGGTGGGTGGGGTTTGCATATTGAAGGTCCAAGC 540
Db       481 TACCTCTATAATCATCAAAACAAGGATGGTGGGTGGGGTTTGCATATTGAAGGTCCAAGC 540

Qy       541 ACCATGTTTGGCTCTGTCTTGAGTTATATTACTCTGAGATTGCTAGGTGAGGGGCCTAAT 600
Db       541 ACCATGTTTGGCTCTGTCTTGAGTTATATTACTCTGAGATTGCTAGGTGAGGGGCCTAAT 600

Qy       601 GATGGACAAGGGGAAATGGAGAAGGCACGTGACTGGATTCTAGGGCATGGTGGTGCTACT 660
Db       601 GATGGACAAGGGGAAATGGAGAAGGCACGTGACTGGATTCTAGGGCATGGTGGTGCTACT 660

Qy       661 TATATAACGTCATGGGGGAAGATGTGGCTTTCAGTACTTGGAGTGTATGAATGGTCTGGA 720
Db       661 TATATAACGTCATGGGGGAAGATGTGGCTTTCAGTACTTGGAGTGTATGAATGGTCTGGA 720

Qy       721 AATAATCCCCTGCCCCCTGAGATATGGCTCCTTCCATACATGCTTCCATTTCATCCAGGA 780
Db       721 AATAATCCCCTGCCCCCTGAGATATGGCTCCTTCCATACATGCTTCCATTTCATCCAGGA 780

Qy       781 AGGATGTGGTGTCACTGCCGGATGGTCTATTTGCCGATGTCCTACTTATATGGCAAGAGG 840
Db       781 AGGATGTGGTGTCACTGCCGGATGGTCTATTTGCCGATGTCCTACTTATATGGCAAGAGG 840

Qy       841 TTTGTTGGTCCAATCTCACCAACAGTATTATCTTTGAGAAAAGAGCTTTATACAGTACCA 900
Db       841 TTTGTTGGTCCAATCTCACCAACAGTATTATCTTTGAGAAAAGAGCTTTATACAGTACCA 900

Qy       901 TACCATGATATAGATTGGGATCAGGCTCGCAATTTGTGTGCAAAGGAAGATTTGTACTAT 960
Db       901 TACCATGATATAGATTGGGATCAGGCTCGCAATTTGTGTGCAAAGGAAGATTTGTACTAT 960

Qy       961 CCTCACCCACTTGTACAGGATATTCTTTGGGCATCTCTACACAAGTTCCTTGAGCCTATT 1020
Db       961 CCTCACCCACTTGTACAGGATATTCTTTGGGCATCTCTACACAAGTTCCTTGAGCCTATT 1020

Qy      1021 CTGATGCATTGGCCTGGAAAAAGATTGAGGGAAAAGGCTATTATTTCTGCATTGGAGCAT 1080
Db      1021 CTGATGCATTGGCCTGGAAAAAGATTGAGGGAAAAGGCTATTATTTCTGCATTGGAGCAT 1080

Qy      1081 ATACATTACGAAGATGAGAATACTCGATATATTTGCATAGGTCCTGTAAATAAGGTGTTA 1140
Db      1081 ATACATTACGAAGATGAGAATACTCGATATATTTGCATAGGTCCTGTAAATAAGGTGTTA 1140

Qy      1141 AATATGCTTTGCTGTTGGGTGGAAGATCCAAATTCTGAGGCCTTCAAGTTGCATCTTCCC 1200
Db      1141 AATATGCTTTGCTGTTGGGTGGAAGATCCAAATTCTGAGGCCTTCAAGTTGCATCTTCCC 1200

Qy      1201 AGGATTTATGATTATCTATGGATTGCAGAAGATGGCATGAAAATGCAGGGCTACAATGGA 1260
Db      1201 AGGATTTATGATTATCTATGGATTGCAGAAGATGGCATGAAAATGCAGGGCTACAATGGA 1260

Qy      1261 AGTCAACTATGGGACACTGCTTTTGCTGTCCAAGCAATTATTGCATCTAACCTCATTGAA 1320
Db      1261 AGTCAACTATGGGACACTGCTTTTGCTGTCCAAGCAATTATTGCATCTAACCTCATTGAA 1320

Qy      1321 GAATTTGGTCCAACTATAAGAAAAGCTCATACCTATATTAAGAATTCACAGGTTTTAGAA 1380
Db      1321 GAATTTGGTCCAACTATAAGAAAAGCTCATACCTATATTAAGAATTCACAGGTTTTAGAA 1380

Qy      1381 GATTGTCCAGGTGATCTTAATAAATGGTACCGTCACATTTCAAAAGGTGCTTGGCCTTTT 1440
Db      1381 GATTGTCCAGGTGATCTTAATAAATGGTACCGTCACATTTCAAAAGGTGCTTGGCCTTTT 1440

Qy      1441 TCAACTGGAGATCATGGATGGCCAATTTCTGACTGCACAGCTGAAGGACTGAAAGCTGTT 1500
Db      1441 TCAACTGGAGATCATGGATGGCCAATTTCTGACTGCACAGCTGAAGGACTGAAAGCTGTT 1500

Qy      1501 CTATTACTATCCAAAATTGCACCAGAAATAGTTGGTGAGCCAATAGACGTGAAGCGATTA 1560
Db      1501 CTATTACTATCCAAAATTGCACCAGAAATAGTTGGTGAGCCAATAGACGTGAAGCGATTA 1560

Qy      1561 TATGATTCTGTAAATGTCATTCTCTCACTACAGAATGAAGATGGTGGTTTTGCAACATAT 1620
Db      1561 TATGATTCTGTAAATGTCATTCTCTCACTACAGAATGAAGATGGTGGTTTTGCAACATAT 1620

Qy      1621 GAGCTTAAACGATCTTATAATTGGTTGGAGATAATCAATCCTGCTGAAACTTTTGGTGAC 1680
Db      1621 GAGCTTAAACGATCTTATAATTGGTTGGAGATAATCAATCCTGCTGAAACTTTTGGTGAC 1680

Qy      1681 ATCGTTATTGATTATCCTTATGTGGAATGTACATCAGCAGCGATTCAAGCTTTGGCATCA 1740
Db      1681 ATCGTTATTGATTATCCTTATGTGGAATGTACATCAGCAGCGATTCAAGCTTTGGCATCA 1740

Qy      1741 TTTAGGAAATTATATCCTGGGCATCGCCGAGAAGAAATACAACATTGTATCGATAAAGCC 1800
Db      1741 TTTAGGAAATTATATCCTGGGCATCGCCGAGAAGAAATACAACATTGTATCGATAAAGCC 1800

Qy      1801 ACTACCTTCATTGAAAAAATACAAGCTTCAGATGGATCATGGTATGGTTCTTGGGGAGTT 1860
Db      1801 ACTACCTTCATTGAAAAAATACAAGCTTCAGATGGATCATGGTATGGTTCTTGGGGAGTT 1860

Qy      1861 TGCTTCACTTACGGTGCTTGGTTTGGGGTAAAAGGTCTGATTGCTGCTGGAAGGAGTTTC 1920
Db      1861 TGCTTCACTTACGGTGCTTGGTTTGGGGTAAAAGGTCTGATTGCTGCTGGAAGGAGTTTC 1920

Qy      1921 AGTAATTGCTCAAGCATCCGTAAAGCTTGTGAATTTCTGCTGTCCAAGCAGCTTCCTTCT 1980
Db      1921 AGTAATTGCTCAAGCATCCGTAAAGCTTGTGAATTTCTGCTGTCCAAGCAGCTTCCTTCT 1980

Qy      1981 GGTGGCTGGGGAGAGAGTTATCTGTCCTGTCAAAACAAGGTGTATTCAAATCTGGAAGGC 2040
Db      1981 GGTGGCTGGGGAGAGAGTTATCTGTCCTGTCAAAACAAGGTGTATTCAAATCTGGAAGGC 2040

Qy      2041 AACAGGTCTCATGTGGTCAACACTGGGTGGGCTATGTTGGCTCTCATTGATGCTGGACAG 2100
Db      2041 AACAGGTCTCATGTGGTCAACACTGGGTGGGCTATGTTGGCTCTCATTGATGCTGGACAG 2100

Qy      2101 GCTAAGAGAGATTCGCAACCACTGCACCGGGCAGCTGCATACTTGATAAATTCCCAATTG 2160
Db      2101 GCTAAGAGAGATTCGCAACCACTGCACCGGGCAGCTGCATACTTGATAAATTCCCAATTG 2160

Qy      2161 GAGGACGGTGACTTTCCGCAGCAGGAAATAATGGGAGTCTTCAACAAGAATTGCATGATC 2220
Db      2161 GAGGACGGTGACTTTCCGCAGCAGGAAATAATGGGAGTCTTCAACAAGAATTGCATGATC 2220

Qy      2221 ACATACGCCGCATACAGAAACATATTCCCCATTTGGGCGTTGGGAGAATACCAATCCCAA 2280
Db      2221 ACATACGCCGCATACAGAAACATATTCCCCATTTGGGCGTTGGGAGAATACCAATCCCAA 2280

Qy      2281 GTATTGCAATCTCGTTAATCGAGCCTTAGTTAGGGTGTCATCCTAATTATTTCGACCTGT 2340
Db      2281 GTATTGCAATCTCGTTAATCGAGCCTTAGTTAGGGTGTCATCCTAATTATTTCGACCTGT 2340

Qy      2341 GGACAGTAAAAGAAATAATAATAATAACCTATTTTTCTTTTATTTCCATGGCTCTCTTAA 2400
Db      2341 GGACAGTAAAAGAAATAATAATAATAACCTATTTTTCTTTTATTTCCATGGCTCTCTTAA 2400

Qy      2401 AATGTTTGTGACTAATGAGTTTAGTAGTCAGCTAAAAAAAAAAAGCAAACACGTGGAGAAT 2460
Db      2401 AATGTTTGTGACTAATGAGTTTAGTAGTCAGCTAAAAAAAAAAGCAAACACGTGGAGAAT 2460

Qy      2461 GCCTGTAAGTTTTTTCTATTACTCATAGACGCCTCTCTTGCTTTCCCTGCAGCAGAGGAA 2520
Db      2461 GCCTGTAAGTTTTTTCTATTACTCATAGACGCCTCTCTTGCTTTCCCTGCAGCAGAGGAA 2520

Qy      2521 TTAAATATATACATAAATAGAGATATAAAAAAAAAAAAAA 2560
Db      2521 TTAAATATATACATAAATAGAGATATAAAAAAAAAAAAAA 2560
```

```
Qy       1981 GGTGGCTGGGGAGAGAGTTATCTGTCCTGTCAAAACAAGGTGTATTCAAATCTGGAAGGC 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 GGTGGCTGGGGAGAGAGTTATCTGTCCTGTCAAAACAAGGTGTATTCAAATCTGGAAGGC 2040

Qy       2041 AACAGGTCTCATGTGGTCAACACTGGGTGGGCTATGTTGGCTCTCATTGATGCTGGACAG 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 AACAGGTCTCATGTGGTCAACACTGGGTGGGCTATGTTGGCTCTCATTGATGCTGGACAG 2100

Qy       2101 GCTAAGAGAGATTCGCAACCACTGCACCGGGCAGCTGCATACTTGATAAATTCCCAATTG 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 GCTAAGAGAGATTCGCAACCACTGCACCGGGCAGCTGCATACTTGATAAATTCCCAATTG 2160

Qy       2161 GAGGACGGTGACTTTCCGCAGCAGGAAATAATGGGAGTCTTCAACAAGAATTGCATGATC 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 GAGGACGGTGACTTTCCGCAGCAGGAAATAATGGGAGTCTTCAACAAGAATTGCATGATC 2220

Qy       2221 ACATACGCCGCATACAGAAACATATTCCCCATTTGGGCGTTGGGAGAATACCAATCCCAA 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 ACATACGCCGCATACAGAAACATATTCCCCATTTGGGCGTTGGGAGAATACCAATCCCAA 2280

Qy       2281 GTATTGCAATCTCGTTAATCGAGCCTTAGTTAGGGTGTCATCCTAATTATTTCGACCTGT 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 GTATTGCAATCTCGTTAATCGAGCCTTAGTTAGGGTGTCATCCTAATTATTTCGACCTGT 2340

Qy       2341 GGACAGTAAAAGAAATAATAATAATAACCTATTTTTCTTTTATTTCCATGGCTCTCTTAA 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GGACAGTAAAAGAAATAATAATAATAACCTATTTTTCTTTTATTTCCATGGCTCTCTTAA 2400

Qy       2401 AATGTTTGTGACTAATGAGTTTAGTAGTCAGCTAAAAAAAAAAAGCAAACACGTGGAGAAT 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 AATGTTTGTGACTAATGAGTTTAGTAGTCAGCTAAAAAAAAAAGCAAACACGTGGAGAAT 2460

Qy       2461 GCCTGTAAGTTTTTTCTATTACTCATAGACGCCTCTCTTGCTTTCCCTGCAGCAGAGGAA 2520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2461 GCCTGTAAGTTTTTTCTATTACTCATAGACGCCTCTCTTGCTTTCCCTGCAGCAGAGGAA 2520

Qy       2521 TTAAATATATACATAAATAGAGATATAAAAAAAAAAAAAA 2560
             ||||||||||||||||||||||||||||||||||||||||
Db       2521 TTAAATATATACATAAATAGAGATATAAAAAAAAAAAAAA 2560
```

RESULT 2
ADQ95187
ID ADQ95187 standard; cDNA; 2560 BP.
XX
AC ADQ95187;
XX
DT 23-SEP-2004 (first entry)
XX
DE Soybean cycloartenol synthase cDNA.
XX
KW soybean; cycloartenol synthase; squalene metabolism; gene; ss.
XX
OS Glycine max.
XX
FH Key Location/Qualifiers
FT CDS 25. .2298
FT /*tag= a
FT /product= "Soybean cycloartenol synthase"
XX
PN US2004121439-A1.
XX
PD 24-JUN-2004.
XX
PF 27-JUN-2003; 2003US-00607726.
XX
PR 23-OCT-1998; 98US-0105405P.
PR 15-OCT-1999; 99US-00419679.
XX
PA (FALC/) FALCO S C.
PA (FAMO/) FAMODU O O.
PA (HITZ/) HITZ W D.
PA (KINN/) KINNEY A J.
PA (RAFA/) RAFALSKI J A.
PA (MCGO/) MCGONIGLE B.
PA (LOHM/) LOHMAN K N.
XX
PI Falco SC, Famodu OO, Hitz WD, Kinney AJ, Rafalski JA;
PI Mcgonigle B, Lohman KN;
XX
DR WPI; 2004-479684/45.
DR P-PSDB; ADQ95188.
XX
PT New composition comprising enzymes involved in squalene metabolism,
PT useful in altering squalene metabolism in plants and seeds.
XX
PS Claim 2; SEQ ID NO 5; 33pp; English.
XX
CC The invention describes a composition comprising an isolated
CC polynucleotide comprising a nucleotide sequence encoding a first
CC polypeptide of at least 200 amino acids. A composition comprising an
CC isolated polynucleotide comprising a nucleotide sequence encoding a first
CC polypeptide of at least 200 amino acids that has at least 90% identity
CC based on the Clustal method of alignment when compared to a polypeptide
CC comprising a sequence of 701, 541, 757, 345, 386, 330, 523, or 467 amino
CC acids (EVEN SEQ ID NOS: 2-16), or an isolated polynucleotide comprising
CC the complement of the nucleotide sequence, or comprising a polypeptide of
CC at least 200 amino acids that has at least 90% identity based on the
CC Clustal method of alignment when compared to a polypeptide of EVEN SEQ ID
CC NOS: 2-16. Also described are: a chimeric gene comprising the isolated
CC polynucleotide described above operably linked to suitable regulatory
CC sequences; an isolated host cell comprising the chimeric gene of (1) or
CC the isolated polynucleotide described above; a virus comprising the
CC isolated polynucleotide described above; selecting an isolated
CC polynucleotide that affects the level of expression of a polypeptide in a
CC host or plant cell; obtaining a nucleic acid fragment encoding a
CC polypeptide; and evaluating at least one compound for its ability to
CC inhibit the activity of a polypeptide. The composition, methods,
CC polynucleotides and polypeptides are useful in altering squalene
CC metabolism in plants and seeds. This sequence encodes soybean
CC cycloartenol synthase, a member of the squalene metabolism pathway.
XX
SQ Sequence 2560 BP; 747 A; 483 C; 596 G; 734 T; 0 U; 0 Other;

Query Match 100.0%; Score 2560; DB 12; Length 2560;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2560; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 TTGGCCTCTTGCCAGCAAAACAGAATGTGGAAGCTCAAGTTCGCCGAAGGAGGGAATCCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TTGGCCTCTTGCCAGCAAAACAGAATGTGGAAGCTCAAGTTCGCCGAAGGAGGGAATCCA 60

Qy         61 TGGCTTCGGACATTGAACAATCACGTTGGAAGACAGGTGTGGGAGTTCGATCCTAAGCTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGGCTTCGGACATTGAACAATCACGTTGGAAGACAGGTGTGGGAGTTCGATCCTAAGCTT 120

Qy        121 GGATCGCCGCAAGATCTCCTCGAGATTGAGAAAGCTCGCCAGAATTTTCACGATAACCGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATCGCCGCAAGATCTCCTCGAGATTGAGAAAGCTCGCCAGAATTTTCACGATAACCGC 180

Qy        181 TTTACCCACAAACACAGCGCTGATCTACTTATGCGGATGCAGTTCGCAAGAGAGAACCCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTTACCCACAAACACAGCGCTGATCTACTTATGCGGATGCAGTTCGCAAGAGAGAACCCA 240

Qy        241 ACACGTGAAGTCTTGCCCAAAGTCGGAGTTAAGGATATTGAGGATGTGACCCAAGAGATT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACACGTGAAGTCTTGCCCAAAGTCGGAGTTAAGGATATTGAGGATGTGACCCAAGAGATT 300

Qy        301 GTGACAAAAACATTAAGAAGGGCCGTAAGTTTCCATTCAACTCTCCAGTGCCATGACGGA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GTGACAAAAACATTAAGAAGGGCCGTAAGTTTCCATTCAACTCTCCAGTGCCATGACGGA 360

Qy        361 CACTGGCCGGGAGATTATGGAGGTCCCATGTTTCTGATGCCTGGCTTGGTAATTACTCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CACTGGCCGGGAGATTATGGAGGTCCCATGTTTCTGATGCCTGGCTTGGTAATTACTCTG 420
```

CC used to create transgenic plants in which the polypeptides are present at
CC higher or lower levels than normal or are found within cell types or
CC developmental stages in which they are not normally found. Furthermore,
CC the polynucleotide may be used as a probe for genetically and physically
CC mapping genes and as a marker for traits linked to those genes, as well
CC as in plant breeding in order to develop lines with desired phenotypes.
CC The current sequence is that of the soybean cycloartenol synthase cDNA of
CC the invention.
XX
SQ Sequence 2560 BP; 747 A; 483 C; 596 G; 734 T; 0 U; 0 Other;

Query Match 100.0%; Score 2560; DB 10; Length 2560;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2560; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 TTGGCCTCTTGCCAGCAAAACAGAATGTGGAAGCTCAAGTTCGCCGAAGGAGGGAATCCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TTGGCCTCTTGCCAGCAAAACAGAATGTGGAAGCTCAAGTTCGCCGAAGGAGGGAATCCA 60

Qy         61 TGGCTTCGGACATTGAACAATCACGTTGGAAGACAGGTGTGGGAGTTCGATCCTAAGCTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGGCTTCGGACATTGAACAATCACGTTGGAAGACAGGTGTGGGAGTTCGATCCTAAGCTT 120

Qy        121 GGATCGCCGCAAGATCTCCTCGAGATTGAGAAAGCTCGCCAGAATTTTCACGATAACCGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATCGCCGCAAGATCTCCTCGAGATTGAGAAAGCTCGCCAGAATTTTCACGATAACCGC 180

Qy        181 TTTACCCACAAACACAGCGCTGATCTACTTATGCGGATGCAGTTCGCAAGAGAGAACCCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTTACCCACAAACACAGCGCTGATCTACTTATGCGGATGCAGTTCGCAAGAGAGAACCCA 240

Qy        241 ACACGTGAAGTCTTGCCCAAAGTCGGAGTTAAGGATATTGAGGATGTGACCCAAGAGATT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACACGTGAAGTCTTGCCCAAAGTCGGAGTTAAGGATATTGAGGATGTGACCCAAGAGATT 300

Qy        301 GTGACAAAAACATTAAGAAGGGCCGTAAGTTTCCATTCAACTCTCCAGTGCCATGACGGA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GTGACAAAAACATTAAGAAGGGCCGTAAGTTTCCATTCAACTCTCCAGTGCCATGACGGA 360

Qy        361 CACTGGCCGGGAGATTATGGAGGTCCCATGTTTCTGATGCCTGGCTTGGTAATTACTCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CACTGGCCGGGAGATTATGGAGGTCCCATGTTTCTGATGCCTGGCTTGGTAATTACTCTG 420

Qy        421 TCTATCACTGGGGCGTTGAATACAGTCTTAACTGAAGAACATAGAAAGGAAATATGCCGT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TCTATCACTGGGGCGTTGAATACAGTCTTAACTGAAGAACATAGAAAGGAAATATGCCGT 480

Qy        481 TACCTCTATAATCATCAAAACAAGGATGGTGGGTGGGGTTTGCATATTGAAGGTCCAAGC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TACCTCTATAATCATCAAAACAAGGATGGTGGGTGGGGTTTGCATATTGAAGGTCCAAGC 540

Qy        541 ACCATGTTTGGCTCTGTCTTGAGTTATATTACTCTGAGATTGCTAGGTGAGGGGCCTAAT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ACCATGTTTGGCTCTGTCTTGAGTTATATTACTCTGAGATTGCTAGGTGAGGGGCCTAAT 600

Qy        601 GATGGACAAGGGGAAATGGAGAAGGCACGTGACTGGATTCTAGGGCATGGTGGTGCTACT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GATGGACAAGGGGAAATGGAGAAGGCACGTGACTGGATTCTAGGGCATGGTGGTGCTACT 660

Qy        661 TATATAACGTCATGGGGGAAGATGTGGCTTTCAGTACTTGGAGTGTATGAATGGTCTGGA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 TATATAACGTCATGGGGGAAGATGTGGCTTTCAGTACTTGGAGTGTATGAATGGTCTGGA 720

Qy        721 AATAATCCCCTGCCCCCTGAGATATGGCTCCTTCCATACATGCTTCCATTTCATCCAGGA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AATAATCCCCTGCCCCCTGAGATATGGCTCCTTCCATACATGCTTCCATTTCATCCAGGA 780

Qy        781 AGGATGTGGTGTCACTGCCGGATGGTCTATTTGCCGATGTCCTACTTATATGGCAAGAGG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGGATGTGGTGTCACTGCCGGATGGTCTATTTGCCGATGTCCTACTTATATGGCAAGAGG 840

Qy        841 TTTGTTGGTCCAATCTCACCAACAGTATTATCTTTGAGAAAAGAGCTTTATACAGTACCA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TTTGTTGGTCCAATCTCACCAACAGTATTATCTTTGAGAAAAGAGCTTTATACAGTACCA 900

Qy        901 TACCATGATATAGATTGGGATCAGGCTCGCAATTTGTGTGCAAAGGAAGATTTGTACTAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TACCATGATATAGATTGGGATCAGGCTCGCAATTTGTGTGCAAAGGAAGATTTGTACTAT 960

Qy        961 CCTCACCCACTTGTACAGGATATTCTTTGGGCATCTCTACACAAGTTCCTTGAGCCTATT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCTCACCCACTTGTACAGGATATTCTTTGGGCATCTCTACACAAGTTCCTTGAGCCTATT 1020

Qy       1021 CTGATGCATTGGCCTGGAAAAAGATTGAGGGAAAAGGCTATTATTTCTGCATTGGAGCAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CTGATGCATTGGCCTGGAAAAAGATTGAGGGAAAAGGCTATTATTTCTGCATTGGAGCAT 1080

Qy       1081 ATACATTACGAAGATGAGAATACTCGATATATTTGCATAGGTCCTGTAAATAAGGTGTTA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ATACATTACGAAGATGAGAATACTCGATATATTTGCATAGGTCCTGTAAATAAGGTGTTA 1140

Qy       1141 AATATGCTTTGCTGTTGGGTGGAAGATCCAAATTCTGAGGCCTTCAAGTTGCATCTTCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AATATGCTTTGCTGTTGGGTGGAAGATCCAAATTCTGAGGCCTTCAAGTTGCATCTTCCC 1200

Qy       1201 AGGATTTATGATTATCTATGGATTGCAGAAGATGGCATGAAAATGCAGGGCTACAATGGA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 AGGATTTATGATTATCTATGGATTGCAGAAGATGGCATGAAAATGCAGGGCTACAATGGA 1260

Qy       1261 AGTCAACTATGGGACACTGCTTTTGCTGTCCAAGCAATTATTGCATCTAACCTCATTGAA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AGTCAACTATGGGACACTGCTTTTGCTGTCCAAGCAATTATTGCATCTAACCTCATTGAA 1320

Qy       1321 GAATTTGGTCCAACTATAAGAAAAGCTCATACCTATATTAAGAATTCACAGGTTTTAGAA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 GAATTTGGTCCAACTATAAGAAAAGCTCATACCTATATTAAGAATTCACAGGTTTTAGAA 1380

Qy       1381 GATTGTCCAGGTGATCTTAATAAATGGTACCGTCACATTTCAAAAGGTGCTTGGCCTTTT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GATTGTCCAGGTGATCTTAATAAATGGTACCGTCACATTTCAAAAGGTGCTTGGCCTTTT 1440

Qy       1441 TCAACTGGAGATCATGGATGGCCAATTTCTGACTGCACAGCTGAAGGACTGAAAGCTGTT 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TCAACTGGAGATCATGGATGGCCAATTTCTGACTGCACAGCTGAAGGACTGAAAGCTGTT 1500

Qy       1501 CTATTACTATCCAAAATTGCACCAGAAATAGTTGGTGAGCCAATAGACGTGAAGCGATTA 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTATTACTATCCAAAATTGCACCAGAAATAGTTGGTGAGCCAATAGACGTGAAGCGATTA 1560

Qy       1561 TATGATTCTGTAAATGTCATTCTCTCACTACAGAATGAAGATGGTGGTTTTGCAACATAT 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TATGATTCTGTAAATGTCATTCTCTCACTACAGAATGAAGATGGTGGTTTTGCAACATAT 1620

Qy       1621 GAGCTTAAACGATCTTATAATTGGTTGGAGATAATCAATCCTGCTGAAACTTTTGGTGAC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 GAGCTTAAACGATCTTATAATTGGTTGGAGATAATCAATCCTGCTGAAACTTTTGGTGAC 1680

Qy       1681 ATCGTTATTGATTATCCTTATGTGGAATGTACATCAGCAGCGATTCAAGCTTTGGCATCA 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 ATCGTTATTGATTATCCTTATGTGGAATGTACATCAGCAGCGATTCAAGCTTTGGCATCA 1740

Qy       1741 TTTAGGAAATTATATCCTGGGCATCGCCGAGAAGAAATACAACATTGTATCGATAAAGCC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 TTTAGGAAATTATATCCTGGGCATCGCCGAGAAGAAATACAACATTGTATCGATAAAGCC 1800

Qy       1801 ACTACCTTCATTGAAAAAATACAAGCTTCAGATGGATCATGGTATGGTTCTTGGGGAGTT 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 ACTACCTTCATTGAAAAAATACAAGCTTCAGATGGATCATGGTATGGTTCTTGGGGAGTT 1860

Qy       1861 TGCTTCACTTACGGTGCTTGGTTTGGGGTAAAAGGTCTGATTGCTGCTGGAAGGAGTTTC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 TGCTTCACTTACGGTGCTTGGTTTGGGGTAAAAGGTCTGATTGCTGCTGGAAGGAGTTTC 1920

Qy       1921 AGTAATTGCTCAAGCATCCGTAAAGCTTGTGAATTTCTGCTGTCCAAGCAGCTTCCTTCT 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 AGTAATTGCTCAAGCATCCGTAAAGCTTGTGAATTTCTGCTGTCCAAGCAGCTTCCTTCT 1980
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2005, 06:24:49 ; Search time 1322 Seconds
(without alignments)
11463.342 Million cell updates/sec

Title: US-09-419-679-5
Perfect score: 2560
Sequence: 1 ttggcctcttgccagcaaaa.....agatataaaaaaaaaaaaaa 2560

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2560	100.0	2560	10	ADE28076	Ade28076 Soybean c
2	2560	100.0	2560	12	ADQ95187	Adq95187 Soybean c
3	1405	54.9	2280	6	ABZ13028	Abz13028 Arabidops
4	1183.6	46.2	2558	10	ADE28072	Ade28072 Corn cycl
5	1183.6	46.2	2558	12	ADQ95183	Adq95183 Corn cycl
6	944.2	36.9	2478	4	AAD16894	Aad16894 Soybean b
7	944.2	36.9	2478	12	ADF89882	Adf89882 G. max be
8	944	36.9	1882	10	ADE28074	Ade28074 Rice cycl
9	944	36.9	1882	12	ADQ95185	Adq95185 Rice cycl
10	905	35.4	2731	4	AAD16893	Aad16893 Soybean o
11	903.8	35.3	2538	4	AAD16896	Aad16896 Wheat oxi
12	897.8	35.1	2766	4	AAD16895	Aad16895 Soybean o
13	897.8	35.1	2766	12	ADF89879	Adf89879 G. max ox
14	847	33.1	2067	10	ADF89798	Adf89798 M. trunca
15	814.8	31.8	2532	3	AAC46759	Aac46759 Arabidops
16	793	31.0	2093	12	ADI45184	Adi45184 Rice isop
17	789.4	30.8	2620	12	ADM28673	Adm28673 DNA encod
18	766.4	29.9	2457	4	AAS09680	Aas09680 Oat cDNA
19	756.4	29.5	2535	12	ADM11044	Adm11044 Ginseng d
20	732	28.6	1546	12	ADI45617	Adi45617 Wheat iso
21	722.2	28.2	2639	12	ADJ48244	Adj48244 Maize oil
22	587.2	22.9	2493	3	AAC42686	Aac42686 Arabidops
23	563.8	22.0	1300	10	ADE28078	Ade28078 Bread whe
24	563.8	22.0	1300	12	ADQ95189	Adq95189 Wheat cyc
25	400	15.6	596	3	AAA69647	Aaa69647 Eucalyptu
26	377	14.7	1013	12	ADF89886	Adf89886 G. max be
27	354.8	13.9	2669	2	AAT59599	Aat59599 Rat-deriv
28	354.8	13.9	2851	10	ADB53186	Adb53186 Primary r
29	343.8	13.4	1356	3	AAC36061	Aac36061 Arabidops
30	337.2	13.2	1358	3	AAC49087	Aac49087 Arabidops
31	328	12.8	2160	10	ADB69885	Adb69885 C. neofor
32	304.8	11.9	3206	6	ABK84268	Abk84268 Human cDN
33	304.8	11.9	3206	10	ADH28966	Adh28966 Human chr
34	304.8	11.9	3206	12	ADP05441	Adp05441 Human lan
35	304.8	11.9	3206	13	ACN40868	Acn40868 Tumour-as
36	304.8	11.9	4346	12	ADE76947	Ade76947 Human cDN
37	303.2	11.8	2637	8	ACA89902	Aca89902 Gene diff
38	280.2	10.9	1926	12	ADI45430	Adi45430 Rice isop
39	280.2	10.9	1926	12	ADJ39566	Adj39566 Plant cDN
40	280.2	10.9	4239	10	ABS56199	Abs56199 cDNA enco
41	280.2	10.9	4239	13	ADS16409	Ads16409 Human lan
42	261.2	10.2	2256	8	ABT19094	Abt19094 Aspergill
43	258	10.1	2208	8	ABT20914	Abt20914 Aspergill
44	249.2	9.7	4480	8	ABX63079	Abx63079 Human cDN
c 45	244.8	9.6	598	12	ADI45527	Adi45527 Banana is

ALIGNMENTS

RESULT 1
ADE28076
ID ADE28076 standard; cDNA; 2560 BP.
XX
AC ADE28076;
XX
DT 29-JAN-2004 (first entry)
XX
DE Soybean cycloartenol synthase cDNA.
XX
KW squalene metabolism; transgenic; plant; genetic; physical gene mapping;
KW breeding; soybean; cycloartenol synthase; ss; gene.
XX
OS Glycine max.
XX
FH Key Location/Qualifiers
FT CDS 25. .2298
FT /*tag= a
FT /product= "Soybean cycloartenol synthase protein"
XX
PN US6630617-B1.
XX
PD 07-OCT-2003.
XX
PF 15-OCT-1999; 99US-00419679.
XX
PR 23-OCT-1998; 98US-0105405P.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Famodu OO, Kinney AJ, Rafalski JA;
XX
DR WPI; 2003-810554/76.
DR P-PSDB; ADE28077.
XX
PT New isolated polynucleotide encoding a squalene metabolic enzyme, useful
PT as probes in gene mapping, in producing transgenic plants or in plant
PT breeding to develop lines having desired phenotypes.
XX
PS Disclosure; SEQ ID NO 5; 31pp; English.
XX
CC The invention relates to a novel isolated polynucleotide encoding a
CC squalene metabolic enzyme. The polynucleotide of the invention may be

```
Qy        560 ValGluCysThrSerAlaAlaIleGlnAlaLeuAlaSerPheArgLysLeuTyrProGly 579
              ||||||||||||||||||||||||||||||||   ||||||:::||||||||||||||||
Db       1681 GTAGAATGTACGTCAGCTGCAATTCAGGCCCTCACATCATTTAAAAAGCTCTATCCTGGG 1740

Qy        580 HisArgArgGluGluIleGlnHisCysIleAspLysAlaThrThrPheIleGluLysIle 599
              |||||||||:::::::::   :::||||||:::||||||      |||||||||   |||
Db       1741 CACCGCAGGAAAGATGTAGATAACTGTATCAACAAAGCTGCTAATTTCATAGAGAGTATT 1800

Qy        600 GlnAlaSerAspGlySerTrpTyrGlySerTrpGlyValCysPheThrTyrGlyAlaTrp 619
              |||   ||||||||||||||||||||||||||   ||||||||||||||||||   |||
Db       1801 CAAAGAAGTGACGGTTCATGGTATGGCTCTTGGGCTGTTTGTTTCACCTACGGCACATGG 1860

Qy        620 PheGlyValLysGlyLeuIleAlaAlaGlyArgSerPheSerAsnCysSerSerIleArg 639
              ||||||||||||   |||:::||||||||||||:::|||   :::      :::||||||
Db       1861 TTTGGAGTGAAGGCATTAGTTGCTGCTGGTAGGACGTTCAAGAGTAGTCCTGCAATCAGG 1920

Qy        640 LysAlaCysGluPheLeuLeuSerLysGlnLeuProSerGlyGlyTrpGlyGluSerTyr 659
              ||||||||||||||||||:::||||||:::||||||   ||||||||||||||||||||
Db       1921 AAGGCATGTGAATTTCTGATGTCAAAAGAGCTTCCTTTCGGTGGCTGGGGAGAAAGTTAT 1980

Qy        660 LeuSerCysGlnAsnLysValTyrSerAsnLeuGluGlyAsnArgSerHisValValAsn 679
              ||||||||||||::::::||||||:::||||||||||||      :::|||   ||||||
Db       1981 TTGTCATGTCAAGATCAGGTTTACACCAATCTTGAAGGCAAGCATGCTCACGCAGTCAAC 2040

Qy        680 ThrGlyTrpAlaMetLeuAlaLeuIleAspAlaGlyGlnAlaLysArgAspSerGlnPro 699
              ||||||||||||||||   ||||||||||||||||||||   ||||||       |||
Db       2041 ACTGGCTGGGCCATGCTGACCTTAATTGACGCAGGACAGGCTGACAGAGATCCAACACCT 2100

Qy        700 LeuHisArgAlaAlaAlaTyrLeuIleAsnSerGlnLeuGluAspGlyAspPheProGln 719
              ||||||||||||||      |||||||||   |||   |||||||||:::|||||||||
Db       2101 CTGCATCGAGCAGCGAAGGTTTTGATCAACTTACAATCTGAGGATGGGGAATTTCCTCAA 2160

Qy        720 GlnGluIleMetGlyValPheAsnLysAsnCysMetIleThrTyrAlaAlaTyrArgAsn 739
              ||||||||||||||||||||||||||||||||||||||:::|||:::   ||||||:::
Db       2161 CAAGAGATCATGGGAGTCTTCAACAAGAACTGCATGATCAGTTACTCCCAGTACAGGGAC 2220

Qy        740 IleProIleTrpAlaLeuGlyGluTyrGlnSerGlnValLeu 754
              |||||||||:::||||||||||||||||||:::   :::||||||
Db       2221 ATATTCCCTGTCTGGGCACTCGGAGAGTACCGTTGCCGAGTCTTG 2265
```

Search completed: November 5, 2005, 15:29:53
Job time : 8608 secs

```
                     /strain="CC7060"
                     /db_xref="taxon:279683"
     mRNA            <1. .>2280
                     /product="cycloartenol synthase"
     CDS             1. .2280
                     /codon_start=1
                     /product="cycloartenol synthase"
                     /protein_id="AAT38890.1"
                     /db_xref="GI:47834385"
                     /translation="MWRLKIAEGGGDPWLRTKNAHVGRQVWEFDPEAGDPEALAAVEA
                     ARRDFAAGRHRLKHSADRLMRIQFEKENPLKSDLPAIKLEENEDVTEEAVSTSLKRAI
                     SRFSTLQAHDGHWPGDYGGPMFLMPGLLITLYVTGSLNTVLSPEHQKEIRRYLYNHQN
                     EDGGWGLHIEGPSTMFGSALTYVSLRLLGEGPESGDGAMEKGRNWILDHGGATYITSW
                     GKFWLAVLGVFDWSGNNPLPPEIWMLPYRLPIHPGRMWCHCRMVYLPMCYVYGKRFVG
                     KITPLILELRSELYKTPYSKIDWDSARNLCAKEDLYYPHPLIQDILWATLHKFVEPVM
                     MHWPGNKLREKALNHVTQHVHYEDENTRYICIGPVNKVLNMLTCWIEDPNSEAPKLHI
                     PRVHDYLWVAEDGMKMQGYNGSQLWDTAFAVQAITATGLIDEFAPTLKLAHNFIKNSQ
                     VLDDCPGDLSYWYRHISKGAWPFSTADHGWPISDCTAEGLKAALLLSKISPEIVGEPV
                     EVNRLYDAVNCLMSWMNNNGGFATYELTRSYAWLELINPAETFGDIVIDYPYVECTSA
                     AIQALTSFKKLYPGHRRKDVDNCINKAANFIESIQRSDGSWYGSWAVCFTYGTWFGVK
                     ALVAAGRTFKSSPAIRKACEFLMSKELPFGGWGESYLSCQDQVYTNLEGKHAHAVNTG
                     WAMLTLIDAGQADRDPTPLHRAAKVLINLQSEDGEFPQQEIMGVFNKNCMISYSQYRD
                     IFPVWALGEYRCRVLAAGK"
ORIGIN
```

Alignment Scores:

Pred. No.:	1.55e-272	Length:	2280
Score:	3265.50	Matches:	568
Percent Similarity:	86.89%	Conservative:	88
Best Local Similarity:	75.23%	Mismatches:	98
Query Match:	78.72%	Indels:	1
DB:	8	Gaps:	1

US-09-419-679-6 (1-757) x AY618692 (1-2280)

```
Qy          1 MetTrpLysLeuLysPheAlaGlu---GlyGlyAsnProTrpLeuArgThrLeuAsnAsn 19
              |||||||:::|||||| ||||||  ||||||:::|||||||||||||||    |||
Db          1 ATGTGGCGGCTGAAGATCGCGGAGGGCGGGGGGCGACCCGTGGCTGCGCACCAAGAACGCG 60

Qy         20 HisValGlyArgGlnValTrpGluPheAspProLysLeuGlySerProGlnAspLeuLeu 39
              ||||||||||||||||||||||||||||||:::   |||   |||:::   |||
Db         61 CACGTCGGCCGCCAGGTCTGGGAGTTCGACCCCGAGGCCGGCGACCCGGAGGCCCTCGCC 120

Qy         40 GluIleGluLysAlaArgGlnAsnPheHisAspAsnArgPheThrHisLysHisSerAla 59
                 :::|||   ||||||::::::|||          |||          ||||||||||
Db        121 GCCGTCGAAGCCGCGCGACGGGACTTCGCCGCGGGCCGCCACCGGCTCAAGCACAGCGCC 180

Qy         60 AspLeuLeuMetArgMetGlnPheAlaArgGluAsnProThrArgGluValLeuProLys 79
              |||   ||||||||:::||||||   :::||||||||||   :::       ||||||
Db        181 GACCGCCTCATGCGCATCCAGTTTGAGAAAGAAAATCCACTTAAGTCGGATCTTCCTGCT 240

Qy         80 ValGlyValLysAspIleGluAspValThrGlnGluIleValThrLysThrLeuArgArg 99
              :::   :::::::::   |||||||||||||:::|||   |||:::   :::|||:::|||
Db        241 ATCAAGCTGGAGGAGAATGAAGATGTCACTGAGGAAGCTGTGTCGACCTCTCTGAAAAGG 300

Qy        100 AlaValSerPheHisSerThrLeuGlnCysHisAspGlyHisTrpProGlyAspTyrGly 119
              |||:::|||       ||||||||||||   ||||||||||||||||||||||||||||
Db        301 GCGATCAGCAGATTTTCTACTCTCCAGGCACACGATGGGCACTGGCCTGGGGATTATGGT 360

Qy        120 GlyProMetPheLeuMetProGlyLeuValIleThrLeuSerIleThrGlyAlaLeuAsn 139
              ||||||||||||||||||||||||||:::|||||||||   :::||||||:::||||||
Db        361 GGTCCTATGTTCCTTATGCCAGGATTGCTTATAACACTATATGTGACTGGATCACTGAAC 420

Qy        140 ThrValLeuThrGluGluHisArgLysGluIleCysArgTyrLeuTyrAsnHisGlnAsn 159
              |||||||||:::   ||||||:::|||||||||   ||||||||||||||||||||||||
Db        421 ACTGTCTTATCACCAGAACATCAGAAGGAGATCCGCCGATATCTCTATAATCATCAGAAT 480

Qy        160 LysAspGlyGlyTrpGlyLeuHisIleGluGlyProSerThrMetPheGlySerValLeu 179
              :::||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GAAGATGGAGGCTGGGGATTGCACATTGAGGGTCCAAGCACCATGTTTGGCTCAGCCTTG 540

Qy        180 SerTyrIleThrLeuArgLeuLeuGlyGluGlyProAsnAspGlyGlnGlyGluMetGlu 199
              :::|||:::::|||||||||||||||||||||||||||      |||  |||  ||||||
```

```
Db        541 ACCTATGTTTCCCTGAGATTGCTTGGAGAGGGACCAGAGAGTGGAGATGGTGCTATGGAA 600

Qy        200 LysAlaArgAspTrpIleLeuGlyHisGlyGlyAlaThrTyrIleThrSerTrpGlyLys 219
              |||   |||:::|||||||||   ||||||||||||||||||||||||||||||||||||
Db        601 AAAGGTCGAAATTGGATTTTAGACCATGGGGGAGCAACTTATATAACATCATGGGGAAAG 660

Qy        220 MetTrpLeuSerValLeuGlyValTyrGluTrpSerGlyAsnAsnProLeuProProGlu 239
              ||||||:::|||||||||||||:::::|||||||||||||||||||||||||||||||||
Db        661 TTTTGGCTCGCGGTACTTGGTGTATTTGATTGGTCTGGTAACAATCCACTGCCACCAGAA 720

Qy        240 IleTrpLeuLeuProTyrMetLeuProPheHisProGlyArgMetTrpCysHisCysArg 259
              |||||:::|||||||||   ||||||   ||||||||||||||||||||||||||||||
Db        721 ATATGGATGTTGCCATACCGCCTGCCAATTCATCCAGGTCGAATGTGGTGTCACTGCCGC 780

Qy        260 MetValTyrLeuProMetSerTyrLeuTyrGlyLysArgPheValGlyProIleSerPro 279
              ||||||||||||||||||   |||:::|||||||||||||||||||||   |||:::|||
Db        781 ATGGTGTATCTACCCATGTGCTATGTTTATGGAAAGAGGTTTGTCGGCAAAATCACACCA 840

Qy        280 ThrValLeuSerLeuArgLysGluLeuTyrThrValProTyrHisAspIleAspTrpAsp 299
              :::|||   ||||||   |||||||||   ||||||   ||||||||||||||||||||
Db        841 CTTATATTGGAATTAAGAAGCGAACTCTACAAAACCCCCTACAGCAAGATTGATTGGGAC 900

Qy        300 GlnAlaArgAsnLeuCysAlaLysGluAspLeuTyrTyrProHisProLeuValGlnAsp 319
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        901 AGTGCTCGCAATCTTTGTGCTAAGGAAGATCTGTACTATCCACATCCATTGATACAGGAT 960

Qy        320 IleLeuTrpAlaSerLeuHisLysPheLeuGluProIleLeuMetHisTrpProGlyLys 339
              ||||||||||||:::|||||||||||:::||||||:::::|||||||||||||||||||
Db        961 ATATTATGGGCCACACTCCATAAATTTGTTGAACCAGTTATGATGCATTGGCCTGGGAAC 1020

Qy        340 ArgLeuArgGluLysAlaIleIleSerAlaLeuGluHisIleHisTyrGluAspGluAsn 359
              :::|||||||||||||||:::            :::|||:::|||||||||||||||||||
Db       1021 AAATTGAGGGAGAAAGCTTTGAACCATGTCACGCAGCATGTTCATTATGAAGATGAGAAC 1080

Qy        360 ThrArgTyrIleCysIleGlyProValAsnLysValLeuAsnMetLeuCysCysTrpVal 379
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db       1081 ACTCGATACATTTGTATTGGTCCTGTGAACAAGGTACTAAATATGCTTACTTGTTGGATT 1140

Qy        380 GluAspProAsnSerGluAlaPheLysLeuHisLeuProArgIleTyrAspTyrLeuTrp 399
              ||||||||||||||||||||||||||||||||||:::||||||::::::||||||||||||
Db       1141 GAAGATCCAAATTCGGAAGCATTCAAGCTCCATATTCCAAGAGTCCATGATTACTTGTGG 1200

Qy        400 IleAlaGluAspGlyMetLysMetGlnGlyTyrAsnGlySerGlnLeuTrpAspThrAla 419
              :::||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GTTGCTGAAGATGGCATGAAGATGCAGGGTTATAATGGCAGCCAGCTCTGGGACACAGCT 1260

Qy        420 PheAlaValGlnAlaIleIleAlaSerAsnLeuIleGluGluPheGlyProThrIleArg 439
              ||||||||||||||||||   |||:::  ||||||:::||||||   ||||||::::::
Db       1261 TTCGCAGTCCAAGCTATTACGGCTACTGGCCTCATTGATGAGTTTGCTCCCACTCTTAAG 1320

Qy        440 LysAlaHisThrTyrIleLysAsnSerGlnValLeuGluAspCysProGlyAspLeuAsn 459
              ||||||   :::|||||||||||||||||||||:::|||||||||||||||||||:::
Db       1321 CTTGCACACAACTTCATTAAGAACTCCCAGGTTCTTGATGACTGCCCTGGAGATCTGAGT 1380

Qy        460 LysTrpTyrArgHisIleSerLysGlyAlaTrpProPheSerThrGlyAspHisGlyTrp 479
              ||||||||||||||||||||||||||||||||||||||||||||||   ||||||||||||
Db       1381 TATTGGTACCGTCACATATCTAAAGGTGCATGGCCTTTTTCTACTGCTGATCATGGTTGG 1440

Qy        480 ProIleSerAspCysThrAlaGluGlyLeuLysAlaValLeuLeuLeuSerLysIleAla 499
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db       1441 CCTATATCAGATTGCACTGCAGAAGGACTAAAGGCCGCATTATTGCTATCGAAGATTTCT 1500

Qy        500 ProGluIleValGlyGluProIleAspValLysArgLeuTyrAspSerValAsnValIle 519
              ||||||||||||||||||:::::|||   ||||||||||:::||||||   :::
Db       1501 CCAGAAATTGTGGGCGAACCAGTGGAAGTTAACAGACTTTATGATGCTGTCAATTGTCTG 1560

Qy        520 LeuSerLeuGlnAsnGluAspGlyGlyPheAlaThrTyrGluLeuLysArgSerTyrAsn 539
              :::|||      |||  :::|||||||||||||||||||||||  ||||||||
Db       1561 ATGTCTTGGATGAACAACAATGGTGGTTTTGCAACTTATGAACTCACAAGGTCTTATGCC 1620

Qy        540 TrpLeuGluIleIleAsnProAlaGluThrPheGlyAspIleValIleAspTyrProTyr 559
              ||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TGGCTAGAGCTTATCAACCCTGCTGAGACCTTTGGGGACATTGTCATTGATTACCCGTAC 1680
```

```
Db        588 GGCCCAAGCACCATGTTTGGTACTGCACTAAATTATATTACTTTGAGATTGCTTGGTGAA 647

Qy        190 GlyProAsnAspGlyGlnGlyGluMetGluLysAlaArgAspTrpIleLeuGlyHisGly 209
              |||   ||||||   |||   :::|||||||||||||   ||||||||   ||||||
Db        648 ---CCTGAGGATGGAATGGGAGCTGTGGAAAAAGCACGGAAGTGGATCTTGGATCATGGT 704

Qy        210 GlyAlaThrTyrIleThrSerTrpGlyLysMetTrpLeuSerValLeuGlyValTyrGlu 229
              |||||||||   ||||||||||||||||||||||||||||||||||||||||||||||||
Db        705 GGTGCAACTGCAATCACTTCATGGGGAAAAATGTGGCTTTCAGTACTTGGAGTGTACGAA 764

Qy        230 TrpSerGlyAsnAsnProLeuProProGluIleTrpLeuLeuProTyrMetLeuProPhe 249
              |||||||||||||||||||||||||||||:::||||||   ||||||:::||||||
Db        765 TGGTCTGGGAATAATCCACTGCCACCAGAAGTATGGCTTTGCCCTTATCTTCTCCCATGT 824

Qy        250 HisProGlyArgMetTrpCysHisCysArgMetValTyrLeuProMetSerTyrLeuTyr 269
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        825 CATCCAGGAAGGATGTGGTGTCATTGCCGAATGGTGTATTTGCCCATGTCATATTTATAT 884

Qy        270 GlyLysArgPheValGlyProIleSerProThrValLeuSerLeuArgLysGluLeuTyr 289
              ||||||||||||||||||||||:::   |||:::   ||||||||||||||||||||||
Db        885 GGTAAGAGGTTTGTTGGTCCTATCACATCCACAATCCAATCTTTGAGAAAGGAGCTTTAT 944

Qy        290 ThrValProTyrHisAspIleAspTrpAspGlnAlaArgAsnLeuCysAlaLysGluAsp 309
              ||||||||||||||:::|||||||||:::::|||||||||   ||||||||||||||||
Db        945 ACAGTCCCTTATCATGAGATCGATTGGAATAAGGCTCGCAATGATTGTGCAAAGGAAGAT 1004

Qy        310 LeuTyrTyrProHisProLeuValGlnAspIleLeuTrpAlaSerLeuHisLysPheLeu 329
              |||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db       1005 CTATACTACCCCCATCCACTGGTTCAGGATATCCTCTGGGCATCTCTTTACTATGCATAT 1064

Qy        330 GluProIleLeuMetHisTrpProGlyLysArgLeuArgGluLysAlaIleIleSerAla 349
              |||||||||   |||:::|||||||   ||||||||||||||||||:::   :::
Db       1065 GAACCCATTTTTATGTATTGGCCTGCAAAAAGATTGAGGGAGAAGGCTCTGGACACTGTG 1124

Qy        350 LeuGluHisIleHisTyrGluAspGluAsnThrArgTyrIleCysIleGlyProValAsn 369
              ::::::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1125 ATGCAGCATATCCACTATGAAGATGAAAATACACGGTATATATGCATAGGGCCTGTCAAT 1184

Qy        370 LysValLeuAsnMetLeuCysCysTrpValGluAspProAsnSerGluAlaPheLysLeu 389
              ||||||||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db       1185 AAGGTGTTAAATATGCTTTGTTGTTGGGCTGAGGATCCAAACTCAGAGGCATTCAAATTA 1244

Qy        390 HisLeuProArgIleTyrAspTyrLeuTrpIleAlaGluAspGlyMetLysMetGlnGly 409
              ||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||
Db       1245 CATCTTCCAAGAATCTTAGATTATTTGTGGATTGCCGAAGATGGCATGAAAATGCAGGGA 1304

Qy        410 TyrAsnGlySerGlnLeuTrpAspThrAlaPheAlaValGlnAlaIleIleAlaSerAsn 429
              ||||||||||||||||||||||||   ||||||||||||||||||||||:::::|||
Db       1305 TATAATGGAAGTCAATTATGGGATACTACTTTTGCTGTTCAAGCGATTATATCAACTAAT 1364

Qy        430 LeuIleGluGluPheGlyProThrIleArgLysAlaHisThrTyrIleLysAsnSerGln 449
              :::   ||||||:::|||   |||:::||||||||||||   |||||||||:::||||||
Db       1365 ATTGCTGAAGAATATGGACAAACTTTAAGAAAAGCACATGAGTACATAAAGGATTCCCAG 1424

Qy        450 ValLeuGluAspCysProGlyAspLeuAsnLysTrpTyrArgHisIleSerLysGlyAla 469
              ||||||||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db       1425 GTCTTAGAAGATTGTCCAGGAGATCTAAATTTTTGGTATCGTCATATTTCAAAGGGTGCT 1484

Qy        470 TrpProPheSerThrGlyAspHisGlyTrpProIleSerAspCysThrAlaGluGlyLeu 489
              ||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||
Db       1485 TGGCCATTCTCAACTGCAGATCATGGATGGCCTATTTCTGACTGTACAGCGGAAGGATTG 1544

Qy        490 LysAlaValLeuLeuLeuSerLysIleAlaProGluIleValGlyGluProIleAspVal 509
              |||||||||:::|||||||||:::       |||   ||||||:::   :::||||||
Db       1545 AAGGCTGTTATATTGCTATCACAATTTCCATCAGAAACAGTTGGGAAATCGGTAGATGTG 1604

Qy        510 LysArgLeuTyrAspSerValAsnValIleLeuSerLeuGlnAsnGluAspGlyGlyPhe 529
              ||||||||||||||:::|||:::||||||||||||||||||   ||||||||||||||
Db       1605 AAGCGGTTATATGATGCTGTACATGTTATCCTTTCCTTACAGAATACTGATGGTGGCTTT 1664

Qy        530 AlaThrTyrGluLeuLysArgSerTyrAsnTrpLeuGluIleIleAsnProAlaGluThr 549
              ||||||||||||||   ||||||||:::|||||||||:::||||||||||||||||||||
Db       1665 GCAACTTATGAATTGACCAGATCCTACCATTGGTTGGAGCTAATCAATCCTGCTGAAACT 1724

Qy        550 PheGlyAspIleValIleAspTyrProTyrValGluCysThrSerAlaAlaIleGlnAla 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1725 TTTGGTGATATTGTCATTGATTACCCATATGTTGAATGTACATCAGCAGCAATTCAAGCA 1784

Qy        570 LeuAlaSerPheArgLysLeuTyrProGlyHisArgArgGluGluIleGlnHisCysIle 589
              |||   |||:::||||||:::|||||||||||||||||||||||:::::||||||
Db       1785 CTGACATTATTCAAGAAATTACATCCTGGGCATAGGAGGGAAGAAATAGAAAATTGTATT 1844

Qy        590 AspLysAlaThrThrPheIleGluLysIleGlnAlaSerAspGlySerTrpTyrGlySer 609
              ||||||   ||||||||||   ||||||||||||||||||||||||||||||||||||
Db       1845 GCTAAAGCTGCAGAGTTTATTGAAAATATACAAGCATCAGATGGCTCTTGGTATGGATCC 1904

Qy        610 TrpGlyValCysPheThrTyrGlyAlaTrpPheGlyValLysGlyLeuIleAlaAlaGly 629
              ||||||||||||||||||||||   ||||||||||:::||||||||||:::||||||||||
Db       1905 TGGGGAGTCTGTTTCACCTATGCTGGCTGGTTTGGGATAAAGGGGTTGGTTGCTGCTGGA 1964

Qy        630 ArgSerPheSerAsnCysSerSerIleArgLysAlaCysGluPheLeuLeuSerLysGln 649
              |||::::::   ||||||||||||||   |||||||||::::::|||||||||||:::
Db       1965 AGGACATACAAGAATTGTTCTAGCATCCACAAGGCATGCGATTATCTGTTGTCCAAAGAG 2024

Qy        650 LeuProSerGlyGlyTrpGlyGluSerTyrLeuSerCysGlnAsnLysValTyrSerAsn 669
              |||   |||||||||||||||||||||||||||||||||||:::|||||||||:::|||
Db       2025 CTCGCTTCAGGTGGTTGGGGAGAAAGTTACCTATCATGTCAGGACAAGGTGTATACCAAT 2084

Qy        670 LeuGluGlyAsnArgSerHisValValAsnThrGlyTrpAlaMetLeuAlaLeuIleAsp 689
              |||:::   ||||||   |||:::|||||||||||||||||||||||||||||||||||
Db       2085 CTCAAAGATAATAGGCCACATATTGTCAATACTGGATGGGCTATGTTGGCCCTCATTGAT 2144

Qy        690 AlaGlyGlnAlaLysArgAspSerGlnProLeuHisArgAlaAlaAlaTyrLeuIleAsn 709
              ||||||||||||:::||||||   ||||||||||||||||||   |||||||||
Db       2145 GCCGGACAGGCTGAGAGAGATCCAACCCCATTGCACCGAGCGGCAAGAATACTAATAAAT 2204

Qy        710 SerGlnLeuGluAspGlyAspPheProGlnGlnGluIleMetGlyValPheAsnLysAsn 729
              ||||||:::|||:::|||||||||||||:::|||||||||||||||||||||||||||||
Db       2205 TCTCAAATGGAGAATGGAGATTTTCCTCAAGAGGAGATTATGGGAGTTTTCAACAAGAAT 2264

Qy        730 CysMetIleThrTyrAlaAlaTyrArgAsnIlePheProIleTrpAlaLeuGlyGluTyr 749
              |||||||||:::|||:::||||||||||||||||||||||||||||||||||||||||||
Db       2265 TGTATGATCAGCTATTCTGCATACCGAAACATTTTTCCTATCTGGGCCCTTGGAGAATAT 2324

Qy        750 GlnSerGlnValLeuGlnSer 756
              :::   :::||||||::::::
Db       2325 CGTTGTCGGGTATTGAAGGCC 2345

RESULT 15
AY618692
LOCUS       AY618692               2280 bp    DNA     linear   PLN 06-JUN-2004
DEFINITION  Avena prostrata strain CC7060 cycloartenol synthase gene, complete
            cds.
ACCESSION   AY618692
VERSION     AY618692.1  GI:47834384
KEYWORDS    .
SOURCE      Avena prostrata
  ORGANISM  Avena prostrata
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Aveneae; Avena.
REFERENCE   1  (bases 1 to 2280)
  AUTHORS   Qi,X., Bakht,S., Leggett,M., Maxwell,C., Melton,R. and Osbourn,A.
  TITLE     A gene cluster for secondary metabolism in oat: Implications for
            the evolution of metabolic diversity in plants
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (21), 8233-8238 (2004)
   PUBMED   15148404
REFERENCE   2  (bases 1 to 2280)
  AUTHORS   Qi,X., Bakht,S., Leggett,M., Maxwell,C., Melton,R. and Osbourn,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAY-2004) Sainsbury Laboratory, John Innes Centre,
            Norwich, Norfolk NR4 7UH, UK
FEATURES             Location/Qualifiers
     source          1. .2280
                     /organism="Avena prostrata"
                     /mol_type="genomic DNA"
```

```
Qy        573 PheArgLysLeuTyrProGlyHisArgArgGluGluIleGlnHisCysIleAspLysAla 592
Db       1743 TTCAAGAAATTATATCCAGGTCATAGAAGGGATGAAATTGATAATTGTGTTGCTAAAGCT 1802

Qy        593 ThrThrPheIleGluLysIleGlnAlaSerAspGlySerTrpTyrGlySerTrpGlyVal 612
Db       1803 GCGGACTTCATTGAAAGCATACAAGCAACTGATGGATCTTGGTATGGATCTTGGGGGGTC 1862

Qy        613 CysPheThrTyrGlyAlaTrpPheGlyValLysGlyLeuIleAlaAlaGlyArgSerPhe 632
Db       1863 TGCTTCACCTATGGTGGTTGGTTTGGGATAAGGGGTCTGGTCGCTGCTGGAAGGAGATAC 1922

Qy        633 SerAsnCysSerSerIleArgLysAlaCysGluPheLeuLeuSerLysGlnLeuProSer 652
Db       1923 GATAACTGTTCTAGCCTTCGTAAAGCTTGTGATTTCCTGTTGTCTAAAGAGCTTGCTTCC 1982

Qy        653 GlyGlyTrpGlyGluSerTyrLeuSerCysGlnAsnLysValTyrSerAsnLeuGluGly 672
Db       1983 GGTGGCTGGGGAGAAAGTTATCTTTCCGGCCAGAATAAGGTGTACACGAATATTAAAGAT 2042

Qy        673 AsnArgSerHisValValAsnThrGlyTrpAlaMetLeuAlaLeuIleAspAlaGlyGln 692
Db       2043 GATAGGCCGCATATTGTCAATACTGGTTGGGCTATGTTATCCCTCATCGATGCTGGGCAG 2102

Qy        693 AlaLysArgAspSerGlnProLeuHisArgAlaAlaAlaTyrLeuIleAsnSerGlnLeu 712
Db       2103 TCTGAGAGAGATCCAACACCGTTACACCGTGCAGCAAGGATATTAATCAATTCTCAGATG 2162

Qy        713 GluAspGlyAspPheProGlnGlnGluIleMetGlyValPheAsnLysAsnCysMetIle 732
Db       2163 GACGATGGAGATTTTCCTCAAGAGGAGATCATGGGAATTTTTAACAAGAACTGCATGATT 2222

Qy        733 ThrTyrAlaAlaTyrArgAsnIlePheProIleTrpAlaLeuGlyGluTyrGlnSerGln 752
Db       2223 AGTTATGCCGCTTATCGTAACATTTTTCCTATATGGGCCCTTGGAGAATATCGCTGCCGG 2282

Qy        753 ValLeuGlnSer 756
Db       2283 GTACTGCAAGCC 2294

RESULT 14
AB055509
LOCUS       AB055509               2460 bp    mRNA    linear   PLN 05-JAN-2002
DEFINITION  Betula platyphylla CASBPX1 mRNA for cycloartenol synthase, complete
            cds.
ACCESSION   AB055509
VERSION     AB055509.1  GI:18147589
KEYWORDS    .
SOURCE      Betula platyphylla
  ORGANISM  Betula platyphylla
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fagales; Betulaceae; Betula.
REFERENCE   1
  AUTHORS   Zhang,H., Shibuya,M., Ebizuka,Y. and Yokota,S.
  TITLE     Oxidosqualene Cyclases from Cell Suspension Culture of Betula
            platyphylla var. japonica:Molecular Evolution of Oxidosqualene
            Cyclases in Higher Plants
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2460)
  AUTHORS   Shibuya,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-FEB-2001) Masaaki Shibuya, University of Tokyo,
            Faculty of Pharmaceutical Sciences; 7-3-1 Hongo, Bunkyo-ku, Tokyo
            113-0033, Japan (E-mail:shibuyam@mol.f.u-tokyo.ac.jp,
            URL:http://www.f.u-tokyo.ac.jp/~tennen/head.htm,
            Tel:81-3-5841-4741, Fax:81-3-5841-4744)
FEATURES             Location/Qualifiers
     source          1. .2460
                     /organism="Betula platyphylla"
                     /mol_type="mRNA"
                     /db_xref="taxon:78630"
     gene            1. .2460
                     /gene="CASBPX1"
     CDS             48. .2351
                     /gene="CASBPX1"
                     /EC_number="5.4.99.8"
                     /codon_start=1
                     /product="cycloartenol synthase"
                     /protein_id="BAB83085.1"
                     /db_xref="GI:18147590"
                     /translation="MWKLKIGAETARGDGGGGGGSETWLRSLNNHLGRQIWEFHPELG
                     TQEELQQIDDARRRFWERRFERRHSSDLLMRIQFAKENPSSANIPQVKIKDTEEVREE
                     AVGMTLRRAINFYSTIQADDGHWPGDYGGPMFLIPGLVITLSITGTLNAFLSKEHQCE
                     ICRYLYNHQNEDGGWGLHIEGPSTMFGTALNYITLRLLGEPEDGMGAVEKARKWILDH
                     GGATAITSWGKMWLSVLGVYEWSGNNPLPPEVWLCPYLLPCHPGRMWCHCRMVYLPMS
                     YLYGKRFVGPITSTIQSLRKELYTVPYHEIDWNKARNDCAKEDLYYPHPLVQDILWAS
                     LYYAYEPIFMYWPAKRLREKALDTVMQHIHYEDENTRYICIGPVNKVLNMLCCWAEDP
                     NSEAFKLHLPRILDYLWIAEDGMKMQGYNGSQLWDTTFAVQAIISTNIAEEYGQTLRK
                     AHEYIKDSQVLEDCPGDLNFWYRHISKGAWPFSTADHGWPISDCTAEGLKAVILLSQF
                     PSETVGKSVDVKRLYDAVHVILSLQNTDGGFATYELTRSYHWLELINPAETFGDIVID
                     YPYVECTSAAIQALTLFKKLHPGHRREEIENCIAKAAEFIENIQASDGSWYGSWGVCF
                     TYAGWFGIKGLVAAGRTYKNCSSIHKACDYLLSKELASGGWGESYLSCQDKVYTNLKD
                     NRPHIVNTGWAMLALIDAGQAERDPTPLHRAARILINSQMENGDFPQEEIMGVFNKNC
                     MISYSAYRNIFPIWALGEYRCRVLKAL"
ORIGIN

Alignment Scores:
Pred. No.:                1.81e-277      Length:          2460
Score:                    3323.00        Matches:         590
Percent Similarity:       88.01%         Conservative:    85
Best Local Similarity:    76.92%         Mismatches:      80
Query Match:              80.11%         Indels:          12
DB:                       8              Gaps:            3

US-09-419-679-6 (1-757) x AB055509 (1-2460)

Qy          1 MetTrpLysLeuLysPhe---AlaGlu----------------------------Gly 9
Db         48 ATGTGGAAGTTGAAGATCGGCGCAGAGACAGCCAGAGGAGATGGTGGTGGCGGCGGCGGC 107

Qy         10 GlyAsnProTrpLeuArgThrLeuAsnAsnHisValGlyArgGlnValTrpGluPheAsp 29
Db        108 AGTGAAACATGGCTGAGGAGTTTGAACAATCACTTGGGACGCCAAATATGGGAGTTCCAT 167

Qy         30 ProLysLeuGlySerProGlnAspLeuLeuGluIleGluLysAlaArgGlnAsnPheHis 49
Db        168 CCAGAGCTCGGCACCCAAGAAGAGCTTCAGCAGATCGACGATGCTCGCCGGAGGTTCTGG 227

Qy         50 AspAsnArgPheThrHisLysHisSerAlaAspLeuLeuMetArgMetGlnPheAlaArg 69
Db        228 GAACGCCGCTTCGAAAGGCGGCACAGCTCCGACCTCCTCATGCGGATTCAGTTTGCAAAG 287

Qy         70 GluAsnProThrArgGluValLeuProLysValGlyValLysAspIleGluAspValThr 89
Db        288 GAGAATCCATCTTCTGCAAATATACCACAAGTCAAAATTAAAGATACAGAGGAGGTCAGA 347

Qy         90 GlnGluIleValThrLysThrLeuArgArgAlaValSerPheHisSerThrLeuGlnCys 109
Db        348 GAGGAGGCTGTAGGAATGACCTTACGAAGGGCCATCAATTTCTACTCAACAATACAGGCA 407

Qy        110 HisAspGlyHisTrpProGlyAspTyrGlyGlyProMetPheLeuMetProGlyLeuVal 129
Db        408 GATGATGGTCATTGGCCTGGGGATTATGGGGGTCCCATGTTTCTAATTCCTGGTCTGGTG 467

Qy        130 IleThrLeuSerIleThrGlyAlaLeuAsnThrValLeuThrGluGluHisArgLysGlu 149
Db        468 ATTACTCTTTCAATTACTGGAACTCTAAATGCTTTCTTATCAAAGGAACATCAATGTGAG 527

Qy        150 IleCysArgTyrLeuTyrAsnHisGlnAsnLysAspGlyGlyTrpGlyLeuHisIleGlu 169
Db        528 ATCTGCAGATATCTCTACAATCATCAGAATGAAGATGGGGGTTGGGGTCTACATATTGAG 587

Qy        170 GlyProSerThrMetPheGlySerValLeuSerTyrIleThrLeuArgLeuLeuGlyGlu 189
```

```
                     /gene="LcCAS1"
                     /codon_start=1
                     /product="cycloartenol synthase"
                     /protein_id="BAA85266.1"
                     /db_xref="GI:6045133"
                     /translation="MWQLKIGADTVPADPSNAGGWLSSLNNHVGRQVWHFHPELGTPE
                     DLQQIQHARQRFSDHREKKHSADLLMRMQFAKNNSSFVNLPQIKVKDKEDVTEEAVS
                     RTLRRAINFYSTIQGDDGHWPGDYGGPMFLIPGLVITLSITGALNAVLSTEHQREICR
                     YLYNHQNKDGGWGLHIEGPSTMFGSVLNYVSLRLLGEEAEDGQGAVDKARKWILDHGG
                     ASAITSWGKMWLSVLGVYEWAGNNPLPPELWLLPYLLPFHPGRMWCHCRMVYLPMCYL
                     YGKRFVGPITPIIRSLRKELYLVPYHEVDWNKARNECAKEDLYYPHPLVQDIVWASLH
                     HVYEPLFMRWPAKRLREKALQCVMQHIHYEDENTRYICIGPVNKVLNMLCCWVEDPHS
                     EAFKLHIPRIFDYLWIAEDGMKMQGYNGSQLWDTAFAVQAIMSTKLAEEYGTTLRKAH
                     KYIKDSQVLEDCPGDLQSWYRHISKGAWPFSTADHGWPISDCTAEGLKAVLLLSKLPS
                     EIVGKSIDEEQIYDAVNVILSLQNTDGGFATYELTRSYPWLELMNPAETFGDIVIDYT
                     YVECTSAAIQALVAFKKLYPGHRRDEIDNCVAKAADFIESIQATDGSWYGSWGVCFTY
                     GGWFGIRGLVAAGRRYDNCSSLRKACDFLLSKELASGGWGESYLSGQNKVYTNIKDDR
                     PHIVNTGWAMLSLIDAGQSERDPTPLHRAARILINSQMDDGDFPQEEIMGIFNKNCMI
                     SYAAYRNIFPIWALGEYRCRVLQAP"
ORIGIN

Alignment Scores:
Pred. No.:                3.61e-278       Length:          2433
Score:                    3331.00         Matches:         586
Percent Similarity:       88.35%          Conservative:    89
Best Local Similarity:    76.70%          Mismatches:      81
Query Match:              80.30%          Indels:          8
DB:                       8               Gaps:            1

US-09-419-679-6 (1-757) x AB033334 (1-2433)

Qy          1 MetTrpLysLeuLysPheAlaGluGlyGlyAsnPro------------------------ 12
              ||||||:::||||||                    |||
Db          3 ATGTGGCAGCTCAAGATTGGGGCCGATACCGTACCTGCCGATCCTTCTAACGCCGGCGGA 62

Qy         13 TrpLeuArgThrLeuAsnAsnHisValGlyArgGlnValTrpGluPheAspProLysLeu 32
              ||||||   :::|||||||||||||||||||||||||||||||   |||   |||:::|||
Db         63 TGGCTCTCATCTCTCAACAACCACGTGGGTCGCCAGGTTTGGCATTTTCATCCCGAACTC 122

Qy         33 GlySerProGlnAspLeuLeuGluIleGluLysAlaArgGlnAsnPheHisAspAsnArg 52
              |||:::|||:::|||||||   :::|||:::   |||||||||   |||   |||:::|||
Db        123 GGCACCCCGGAAGATCTTCAGCAGATACAACATGCGCGCCAGCGCTTCTCTGACCACAGA 182

Qy         53 PheThrHisLysHisSerAlaAspLeuLeuMetArgMetGlnPheAlaArgGluAsnPro 72
              |||      ||||||||||||||||||||||||||||||||||||:::   |||
Db        183 TTCGAGAAGAAGCATAGTGCTGATCTCCTTATGCGGATGCAGTTTGCGAAGAATAACTCA 242

Qy         73 ThrArgGluValLeuProLysValGlyValLysAspIleGluAspValThrGlnGluIle 92
              :::      |||||:::::   |||||||||   |||||||||||:::|||
Db        243 TCCTTTGTGAATCTACCGCAAATCAAAGTTAAAGATAAAGAAGACGTGACAGAGGAGGCA 302

Qy         93 ValThrLysThrLeuArgArgAlaValSerPheHisSerThrLeuGlnCysHisAspGly 112
              |||::::::|||||||||||||:::::|||:::|||||||:::|||      |||||||
Db        303 GTATCGCGCACTCTAAGAAGGGCTATCAATTTTTATTCAACAATCCAGGGAGATGATGGT 362

Qy        113 HisTrpProGlyAspTyrGlyGlyProMetPheLeuMetProGlyLeuValIleThrLeu 132
              ||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        363 CACTGGCCTGGAGATTATGGCGGTCCAATGTTTCTAATTCCTGGTTTGGTCATTACTCTT 422

Qy        133 SerIleThrGlyAlaLeuAsnThrValLeuThrGluGluHisArgLysGluIleCysArg 152
              ||||||||||||||||||   ||||||:::   ||||||:::::||||||||||||
Db        423 TCCATTACTGGTGCCTTGAATGCCGTCTTATCTACAGAACACCAGCGTGAGATTTGCAGA 482

Qy        153 TyrLeuTyrAsnHisGlnAsnLysAspGlyGlyTrpGlyLeuHisIleGluGlyProSer 172
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        483 TATCTATACAATCATCAGAATAAAGATGGGGGTTGGGGTTTACATATTGAAGGCCCAAGC 542

Qy        173 ThrMetPheGlySerValLeuSerTyrIleThrLeuArgLeuLeuGlyGluGlyProAsn 192
              ||||||||||||||||||||:::|||:::::||||||||||||||||||
Db        543 ACAATGTTCGGTTCTGTGCTGAATTATGTTTCTTTGAGATTGCTTGGTGAAGAAGCTGAA 602

Qy        193 AspGlyGlnGlyGluMetGluLysAlaArgAspTrpIleLeuGlyHisGlyGlyAlaThr 212
              ||||||||||||   ::::::|||||||||   |||||||||   ||||||||||||:::

Db        603 GATGGGCAGGGTGCTGTGGACAAAGCACGGAAATGGATCTTGGACCATGGTGGTGCATCT 662

Qy        213 TyrIleThrSerTrpGlyLysMetTrpLeuSerValLeuGlyValTyrGluTrpSerGly 232
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:::|||
Db        663 GCAATTACTTCTTGGGGGAAAATGTGGCTTTCAGTACTTGGAGTATATGAATGGGCTGGA 722

Qy        233 AsnAsnProLeuProProGluIleTrpLeuLeuProTyrMetLeuProPheHisProGly 252
              ||||||||||||||||||||:::|||||||||||||||:::|||||||||||||||||||
Db        723 AATAATCCACTTCCTCCCGAACTATGGCTCTTGCCTTATCTTCTCCCTTTTCATCCAGGA 782

Qy        253 ArgMetTrpCysHisCysArgMetValTyrLeuProMetSerTyrLeuTyrGlyLysArg 272
              ||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db        783 AGGATGTGGTGTCACTGCCGAATGGTATACTTGCCTATGTGTTATCTGTATGGCAAGAGA 842

Qy        273 PheValGlyProIleSerProThrValLeuSerLeuArgLysGluLeuTyrThrValPro 292
              ||||||||||||||||:::|||   :::   ||||||||||||||||||||   ||||||
Db        843 TTTGTTGGTCCTATAACGCCTATAATTCGATCTTTAAGGAAGGAACTTTATCTTGTCCCT 902

Qy        293 TyrHisAspIleAspTrpAspGlnAlaArgAsnLeuCysAlaLysGluAspLeuTyrTyr 312
              |||||:::::||||||::::::||||||||   ||||||||||||||||||||||||||
Db        903 TATCATGAAGTTGATTGGAATAAGGCTCGCAATGAGTGTGCAAAGGAAGATCTGTATTAC 962

Qy        313 ProHisProLeuValGlnAspIleLeuTrpAlaSerLeuHisLysPheLeuGluProIle 332
              ||||||||||||||||||||||:::|||||||||||||||         ||||||:::
Db        963 CCACATCCGCTGGTTCAGGATATTGTGTGGGCTTCGCTTCATCATGTCTATGAGCCTCTT 1022

Qy        333 LeuMetHisTrpProGlyLysArgLeuArgGluLysAlaIleIleSerAlaLeuGluHis 352
              |||   ||||||   ||||||||||||||||||||:::          ::::::|||
Db       1023 TTTATGCGTTGGCCTGCAAAAAGACTGAGGGAAAAGGCTTTGCAGTGTGTAATGCAGCAT 1082

Qy        353 IleHisTyrGluAspGluAsnThrArgTyrIleCysIleGlyProValAsnLysValLeu 372
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1083 ATCCACTATGAAGATGAAAATACTCGGTATATATGCATAGGGCCTGTCAACAAGGTATTA 1142

Qy        373 AsnMetLeuCysCysTrpValGluAspProAsnSerGluAlaPheLysLeuHisLeuPro 392
              ||||||||||||||||||||||||||||||:::|||||||||||||||||||:::|||
Db       1143 AACATGCTTTGTTGTTGGGTGGAGGATCCACATTCGGAGGCATTCAAATTGCATATTCCA 1202

Qy        393 ArgIleTyrAspTyrLeuTrpIleAlaGluAspGlyMetLysMetGlnGlyTyrAsnGly 412
              ||||||:::||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1203 AGAATCTTTGACTATCTGTGGATTGCTGAAGATGGCATGAAAATGCAGGGTTATAATGGA 1262

Qy        413 SerGlnLeuTrpAspThrAlaPheAlaValGlnAlaIleIleAlaSerAsnLeuIleGlu 432
              ||||||||||||||||||||||||||||||||||||||||:::::::::   |||   |||
Db       1263 AGTCAATTATGGGATACTGCATTTGCTGTTCAAGCCATCATGTCAACTAAGCTTGCTGAA 1322

Qy        433 GluPheGlyProThrIleArgLysAlaHisThrTyrIleLysAsnSerGlnValLeuGlu 452
              |||:::|||   |||:::|||||||||||   ||||||||:::|||||||||||||||
Db       1323 GAATATGGGACAACTTTAAGAAAGGCACACAAGTATATAAAAGATTCCCAGGTCTTAGAA 1382

Qy        453 AspCysProGlyAspLeuAsnLysTrpTyrArgHisIleSerLysGlyAlaTrpProPhe 472
              ||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||
Db       1383 GATTGTCCTGGTGATTTACAAAGTTGGTATCGTCATATTTCAAAAGGTGCATGGCCATTT 1442

Qy        473 SerThrGlyAspHisGlyTrpProIleSerAspCysThrAlaGluGlyLeuLysAlaVal 492
              ||||||   ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1443 TCAACTGCAGATCATGGATGGCCCATCTCTGACTGCACTGCTGAAGGTTTGAAGGCTGTT 1502

Qy        493 LeuLeuLeuSerLysIleAlaProGluIleValGlyGluProIleAspValLysArgLeu 512
              ||||||||||||||:::   ||||||||||||:::   ||||||   :::::::::
Db       1503 CTTTTATTATCAAAACTTCCTTCAGAAATAGTAGGGAAGTCAATTGATGAAGAACAGATA 1562

Qy        513 TyrAspSerValAsnValIleLeuSerLeuGlnAsnGluAspGlyGlyPheAlaThrTyr 532
              ||||||:::||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1563 TACGATGCTGTAAATGTCATCCTCTCCTTACAGAATACTGATGGTGGCTTTGCCACTTAC 1622

Qy        533 GluLeuLysArgSerTyrAsnTrpLeuGluIleIleAsnProAlaGluThrPheGlyAsp 552
              ||||||   |||||||||   |||||||||::::::||||||||||||||||||||||||
Db       1623 GAACTGACCAGATCTTACCCGTGGCTGGAGCTAATGAACCCTGCTGAAACTTTCGGTGAT 1682

Qy        553 IleValIleAspTyrProTyrValGluCysThrSerAlaAlaIleGlnAlaLeuAlaSer 572
              ||||||||||||||   ||||||||||||||||||||||||||||||||||||   :::
Db       1683 ATTGTCATTGACTACACATATGTCGAGTGTACCTCAGCAGCAATTCAAGCACTTGTAGCG 1742
```

```
Db        807 GGGCGAAAGTGGATCCTAGATCATGGCAGTGCAACTGCTATAACCTCATGGGGAAAGATG 866

Qy        221 TrpLeuSerValLeuGlyValTyrGluTrpSerGlyAsnAsnProLeuProProGluIle 240

Db        867 TGGTTATCTGTGCTTGGAGTGTTTGACTGGTCTGGCAACAATCCTTTGCCGCCAGAAATG 926

Qy        241 TrpLeuLeuProTyrMetLeuProPheHisProGlyArgMetTrpCysHisCysArgMet 260

Db        927 TGGTTATTACCATACTTCTTGCCAGTTCATCCAGGACGAATGTGGTGTCATTGCCGGATG 986

Qy        261 ValTyrLeuProMetSerTyrLeuTyrGlyLysArgPheValGlyProIleSerProThr 280

Db        987 GTTTACCTGCCCATGTCATACATCTATGGAAAGAGGTTTACTGGTCCAATTACACCACTT 1046

Qy        281 ValLeuSerLeuArgLysGluLeuTyrThrValProTyrHisAspIleAspTrpAspGln 300

Db       1047 ATCTTGTCACTAAGGAAGGAACTATTTAATCTTCCATTTGATCAGTTAGATTGGAACAAG 1106

Qy        301 AlaArgAsnLeuCysAlaLysGluAspLeuTyrTyrProHisProLeuValGlnAspIle 320

Db       1107 GCTCGTAACGAGTGTGCAAAGGAAGATCTGTACTATCCACATCCATTTATACAAGACGTC 1166

Qy        321 LeuTrpAlaSerLeuHisLysPheLeuGluProIleLeuMetHisTrpProGlyLysArg 340

Db       1167 CTTTGGGCATCTCTACATAAATTTGTCGAACCTATTCTGATGCATTGGCCTGGAAGCAAA 1226

Qy        341 LeuArgGluLysAlaIleIleSerAlaLeuGluHisIleHisTyrGluAspGluAsnThr 360

Db       1227 TTAAGAGAGAAGGCTGTTAATACTGCTATGCAGCATGTCCATTATGAGGATGAAAATACA 1286

Qy        361 ArgTyrIleCysIleGlyProValAsnLysValLeuAsnMetLeuCysCysTrpValGlu 380

Db       1287 AGATATATTTGCATTGGTCCTGTCAACAAGGTACTTAATATGCTTTGCTGCTGGATTGAA 1346

Qy        381 AspProAsnSerGluAlaPheLysLeuHisLeuProArgIleTyrAspTyrLeuTrpIle 400

Db       1347 GATCCAAATTCAGAGGCATTCAAGCTACATCTTCCTAGAGTAATGGATTATCTGTGGTTG 1406

Qy        401 AlaGluAspGlyMetLysMetGlnGlyTyrAsnGlySerGlnLeuTrpAspThrAlaPhe 420

Db       1407 GCTGAAGATGGAATGAAAATGCGGGGCTATAATGGCAGCCAGTTGTGGGACACAGCTTTC 1466

Qy        421 AlaValGlnAlaIleIleAlaSerAsnLeuIleGluGluPheGlyProThrIleArgLys 440

Db       1467 ACTGTTCAGGCAATAATATCGACTGACCTTTTTGAGGAATTTGGTCTTGCACTGACGAAG 1526

Qy        441 AlaHisThrTyrIleLysAsnSerGlnValLeuGluAspCysProGlyAspLeuAsnLys 460

Db       1527 GCACATGAATTTATTAAAAAAACACAGGTTCTTGAAGACTGCCCTGGCGATCTTAATTTT 1586

Qy        461 TrpTyrArgHisIleSerLysGlyAlaTrpProPheSerThrGlyAspHisGlyTrpPro 480

Db       1587 TGGTACCGCCACATCTCTAAAGGTGCATGGCCTTTCTCCACCGCTGATCATGGATGGCCT 1646

Qy        481 IleSerAspCysThrAlaGluGlyLeuLysAlaValLeuLeuLeuSerLysIleAlaPro 500

Db       1647 ATCTCAGATTGCACTTCCGAAGGACTTAAGGCAGCGCTCTTATTATCTAAGATTTCTCCT 1706

Qy        501 GluIleValGlyGluProIleAspValLysArgLeuTyrAspSerValAsnValIleLeu 520

Db       1707 GAAATTGTTGGTGATCCACTTGATGGAAAGAGCCTCTATGATGCCGTTAATGTTATCCTT 1766

Qy        521 SerLeuGlnAsnGluAspGlyGlyPheAlaThrTyrGluLeuLysArgSerTyrAsnTrp 540

Db       1767 TCACTAATGAATAATGATGGTGGTTTTGCTACGTATGAACTTACGAGATCCTATGCCTGG 1826

Qy        541 LeuGluIleIleAsnProAlaGluThrPheGlyAspIleValIleAspTyrProTyrVal 560

Db       1827 TTGGAGATTATCAATCCTGCTGAAACATTTGGAGACATTGTTATCGATTACCCATATGTT 1886

Qy        561 GluCysThrSerAlaAlaIleGlnAlaLeuAlaSerPheArgLysLeuTyrProGlyHis 580

Db       1887 GAGTGCACTTCGGCAGCAATTCAGGCATTGACATTATTTAAAAAGACATATCCTGGACAT 1946

Qy        581 ArgArgGluGluIleGlnHisCysIleAspLysAlaThrThrPheIleGluLysIleGln 600

Db       1947 CGCCGAGAGGAGATAGACAATTGCATCAGAAAATCAGCTCGTTTTATTGAAAAGATTCAG 2006

Qy        601 AlaSerAspGlySerTrpTyrGlySerTrpGlyValCysPheThrTyrGlyAlaTrpPhe 620

Db       2007 CTAGCTGATGGTTCATGGTATGGATCTTGGGGTGTTTGCTTCACATATGGTATATGGTTT 2066

Qy        621 GlyValLysGlyLeuIleAlaAlaGlyArgSerPheSerAsnCysSerSerIleArgLys 640

Db       2067 GGAATGAAGGGATTACTTGCCGCTGGTCGGACCTATGAGACTAGCTCTTGCATCCGAAAG 2126

Qy        641 AlaCysGluPheLeuLeuSerLysGlnLeuProSerGlyGlyTrpGlyGluSerTyrLeu 660

Db       2127 GCATGTGACTTTCTGTTGTCTAAACAAGTTGCCTCTGGTGGTTGGGGGGAGAGCTATCTT 2186

Qy        661 SerCysGlnAsnLysValTyrSerAsnLeuGluGlyAsnArgSerHisValValAsnThr 680

Db       2187 TCATGTCAAAATAAGGTCTACACAAATCTTGAGGGTAACAGGGCTCATGCAGTGAATACT 2246

Qy        681 GlyTrpAlaMetLeuAlaLeuIleAspAlaGlyGlnAlaLysArgAspSerGlnProLeu 700

Db       2247 GGCTGGGCCATGCTGGCTCTAATTGATGCTGGGCAGGGTGAGAGAGATCCAAAACCATTG 2306

Qy        701 HisArgAlaAlaAlaTyrLeuIleAsnSerGlnLeuGluAspGlyAspProProGlnGln 720

Db       2307 CACAGAGCAGCAAAAGTTCTAATAAACATGCAAATGGAGAATGGCGAGTTTCCCCAACAA 2366

Qy        721 GluIleMetGlyValPheAsnLysAsnCysMetIleThrTyrAlaAlaTyrArgAsnIle 740

Db       2367 GAGATTATGGGGGTTTTTAACAAAAATTGCATGATCAGTTATTCAGAATATCGAAACATA 2426

Qy        741 ProProIleTrpAlaLeuGlyGluTyrGlnSerGlnValLeuGlnSerArg 757

Db       2427 TTTCCAATCTGGGCTCTCGGAGAATACCGAAGACGCGTCCTGTGTTCTCAA 2477
```

```
RESULT 13
AB033334
LOCUS       AB033334               2433 bp    mRNA    linear   PLN 07-JAN-2000
DEFINITION  Luffa cylindrica LcCAS1 mRNA for cycloartenol synthase, complete
            cds.
ACCESSION   AB033334
VERSION     AB033334.1  GI:6045132
KEYWORDS    cycloartenol synthase.
SOURCE      Luffa cylindrica (smooth loofah)
  ORGANISM  Luffa cylindrica
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Cucurbitales; Cucurbitaceae; Luffa.
REFERENCE   1  (sites)
  AUTHORS   Hayashi,H., Hiraoka,N., Ikeshiro,Y., Yazaki,K., Tanaka,S.,
            Kushiro,T., Shibuya,M. and Ebizuka,Y.
  TITLE     Molecular cloning of a cDNA encoding cycloartenol synthase from
            Luffa cylindrica
  JOURNAL   Plant Physiol. 121, 1383 (1999)
REFERENCE   2  (bases 1 to 2433)
  AUTHORS   Hayashi,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-1999) Hiroaki Hayashi, Gifu Pharmaceutical
            University, Department of Pharmadognosy; 5-6-1 Mitahora-higashi,
            Gifu, Gifu 502-8585, Japan (E-mail:hayashih@gifu-pu.ac.jp,
            Tel:81-58-237-3931(ex.325), Fax:81-58-237-5979)
FEATURES             Location/Qualifiers
     source          1. .2433
                     /organism="Luffa cylindrica"
                     /mol_type="mRNA"
                     /db_xref="taxon:3670"
                     /cell_type="suspension cultured cells"
                     /clone_lib="lambda ZAPII"
     gene            1. .2433
                     /gene="LcCAS1"
     CDS             3. .2300
```

```
Db       1983 TGCTTCACCTATGGTGGTTGGTTCGGGATAAGGGGGCTGGTCGCTGCTGGAAGGAGATAC 2042

Qy        633 SerAsnCysSerSerIleArgLysAlaCysGluPheLeuLeuSerLysGlnLeuProSer 652

Db       2043 AATAACTGTTCTAGCCTTCGTAAAGCTTGTGATTTCCTGTTGTCCAAAGAGCTTGCTGCC 2102

Qy        653 GlyGlyTrpGlyGluSerTyrLeuSerCysGlnAsnLysValTyrSerAsnLeuGluGly 672

Db       2103 GGTGGCTGGGGAGAAAGTTACCTTTCCTGCCAGAATAAGGTGTACACAAATATTAAAGAT 2162

Qy        673 AsnArgSerHisValValAsnThrGlyTrpAlaMetLeuAlaLeuIleAspAlaGlyGln 692

Db       2163 GACAGGCCACATATTGTCAATACTGGTTGGGCTATGTTATCCCTCATCGATGCTGGGCAG 2222

Qy        693 AlaLysArgAspSerGlnProLeuHisArgAlaAlaAlaTyrLeuIleAsnSerGlnLeu 712

Db       2223 TCTGAGAGAGATCCAACACCATTACACCGTGCAGCAAGGGTATTAATTAATTCTCAGATG 2282

Qy        713 GluAspGlyAspPheProGlnGlnGluIleMetGlyValPheAsnLysAsnCysMetIle 732

Db       2283 GAAGACGGAGATTTTCCTCAAGAGGAGATCATGGGAGTCTTTAACAAGAACTGCATGATT 2342

Qy        733 ThrTyrAlaAlaTyrArgAsnIlePheProIleTrpAlaLeuGlyGluTyrGlnSerGln 752

Db       2343 AGCTATTCAGCATACCGTAACATTTTCCCTATATGGGCCCTTGGGGAATACCGTAGCCGG 2402

Qy        753 ValLeuGln 755

Db       2403 GTACTGAAA 2411
```

```
RESULT 12
AB058507
LOCUS       AB058507                2683 bp    mRNA    linear   PLN 10-APR-2002
DEFINITION  Costus speciosus CSOSC1 mRNA for cycloartenol synthase, complete
            cds.
ACCESSION   AB058507
VERSION     AB058507.1  GI:18147770
KEYWORDS    .
SOURCE      Costus speciosus
  ORGANISM  Costus speciosus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Zingiberales; Costaceae;
            Costus.
REFERENCE   1
  AUTHORS   Kawano,N., Ichinose,K. and Ebizuka,Y.
  TITLE     Molecular cloning and functional expression of cDNAs encoding
            oxidosqualene cyclases from Costus speciosus
  JOURNAL   Biol. Pharm. Bull. 25 (4), 477-482 (2002)
  MEDLINE   21990915
   PUBMED   11995928
REFERENCE   2  (bases 1 to 2683)
  AUTHORS   Kawano,N., Ichinose,K. and Ebizuka,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAR-2001) Noriaki Kawano, The University of Tokyo,
            Graduate School of Pharmaceutical Sciences; 7-3-1 Hongo, Bunkyo-ku,
            Tokyo 113-0033, Japan (E-mail:ff17004@mail.ecc.u-tokyo.ac.jp,
            Tel:81-3-5841-4742, Fax:81-3-5841-4744)
FEATURES             Location/Qualifiers
     source          1. .2683
                     /organism="Costus speciosus"
                     /mol_type="mRNA"
                     /db_xref="taxon:49577"
                     /clone="CSOSC1"
                     /tissue_type="whole plantlet"
                     /dev_stage="plantlet"
                     /note="in vitro cultured plantlet"
     gene            1. .2683
                     /gene="CSOSC1"
     CDS             207. .2486
                     /gene="CSOSC1"
                     /EC_number="5.4.99.8"
                     /note="functional expression in Saccharomyces cerevisiae"
                     /codon_start=1
                     /evidence=experimental
                     /product="cycloartenol synthase"
                     /protein_id="BAB83253.1"
                     /db_xref="GI:18147771"
                     /translation="MWRLKIAEGGGPWLRTKNNHVGRQVWEFDPSLGTPEEIAEVERV
                     REAFRETRFEKKHSADLLMRLQFAKENPLEMNYPIIKIEEHEDVTEELVVTSLRKAIS
                     RVSTLQAHDGHWPGDYGGPMFLMPGLIITLYVTGALNTVLTSEHQKEIRRYLYNHQNE
                     DGGWGLHIEGESTMFGSALTYVILRLLGEGPDDGDGAMEKGRKWILDHGSATAITSWG
                     KMWLSVLGVFDWSGNNPLPPEMWLLPYFLPVHPGRMWCHCRMVYLPMSYIYGKRFTGP
                     ITPLILSLRKELFNLPFDQLDWNKARNECAKEDLYYPHPFIQDVLWASLHKFVEPILM
                     HWPGSKLREKAVNTAMQHVHYEDENTRYICIGPVNKVLNMLCCWIEDPNSEAFKLHLP
                     RVMDYLWLAEDGMKMRGYNGSQLWDTAFTVQAIISTDLFEEFGLALTKAHEFIKKTQV
                     LEDCPGDLNFWYRHISKGAWPFSTADHGWPISDCTSEGLKAALLLSKISPEIVGDPLD
                     GKSLYDAVNVILSLMNNDGGFATYELTRSYAWLEIINPAETFGDIVIDYPYVECTSAA
                     IQALTLFKKTYPGHRREEIDNCIRKSARFIEKIQLADGSWYGSWGVCFTYGIWFGMKG
                     LLAAGRTYETSSCIRKACDFLLSKQVASGGWGESYLSCQNKVYTNLEGNRAHAVNTGW
                     AMLALIDAGQGERDPKPLHRAAKVLINMQMENGEFPQQEIMGVFNKNCMISYSEYRNI
                     FPIWALGEYRRRVLCSQFH"
ORIGIN
```

```
Alignment Scores:
Pred. No.:                6.25e-280       Length:          2683
Score:                    3352.00         Matches:         584
Percent Similarity:       88.38%          Conservative:    85
Best Local Similarity:    77.15%          Mismatches:      88
Query Match:              80.81%          Indels:          0
DB:                       8               Gaps:            0

US-09-419-679-6 (1-757) x AB058507 (1-2683)

Qy          1 MetTrpLysLeuLysPheAlaGluGlyGlyAsnProTrpLeuArgThrLeuAsnAsnHis 20

Db        207 ATGTGGAGGCTCAAGATCGCGGAGGGCGGCGGCCCGTGGCTGCGGACGAAGAACAACCAC 266

Qy         21 ValGlyArgGlnValTrpGluPheAspProLysLeuGlySerProGlnAspLeuLeuGlu 40

Db        267 GTCGGGAGGCAGGTGTGGGAGTTCGACCCGAGCCTCGGCACGCCCGAGGAGATAGCGGAG 326

Qy         41 IleGluLysAlaArgGlnAsnPheHisAspAsnArgPheThrHisLysHisSerAlaAsp 60

Db        327 GTGGAGAGAGTCCGGGAGGCTTTCCGCGAGACCCGGTTCGAGAAGAAGCACAGCGCCGAT 386

Qy         61 LeuLeuMetArgMetGlnPheAlaArgGluAsnProThrArgGluValLeuProLysVal 80

Db        387 CTCCTCATGCGGTTGCAGTTTGCCAAGGAGAATCCACTTGAAATGAACTATCCTATCATC 446

Qy         81 GlyValLysAspIleGluAspValThrGlnGluIleValThrLysThrLeuArgArgAla 100

Db        447 AAGATTGAAGAACACGAAGATGTCACTGAAGAACTTGTGGTGACCTCCTTAAGAAAGGCT 506

Qy        101 ValSerPheHisSerThrLeuGlnCysHisAspGlyHisTrpProGlyAspTyrGlyGly 120

Db        507 ATCAGTCGCGTTTCTACTCTCCAAGCTCACGATGGACACTGGCCTGGGGATTATGGTGGT 566

Qy        121 ProMetPheLeuMetProGlyLeuValIleThrLeuSerIleThrGlyAlaLeuAsnThr 140

Db        567 CCAATGTTCCTTATGCCGGGATTGATAATAACTTTGTATGTCACTGGAGCTCTAAACACT 626

Qy        141 ValLeuThrGluGluHisArgLysGluIleCysArgTyrLeuTyrAsnHisGlnAsnLys 160

Db        627 GTTCTTACATCTGAACACCAGAAGGAGATCCGTCGATACCTCTACAACCACCAGAATGAA 686

Qy        161 AspGlyGlyTrpGlyLeuHisIleGluGlyProSerThrMetPheGlySerValLeuSer 180

Db        687 GACGGAGGCTGGGGCTTGCATATTGAGGGTGAAAGCACTATGTTTGGTTCTGCCTTGACT 746

Qy        181 TyrIleThrLeuArgLeuLeuGlyGluGlyProAsnAspGlyGlnGlyGluMetGluLys 200

Db        747 TATGTTATTTTGAGATTGCTCGGAGAAGGACCTGATGATGGAGATGGGGCAATGGAAAAA 806

Qy        201 AlaArgAspTrpIleLeuGlyHisGlyGlyAlaThrTyrIleThrSerTrpGlyLysMet 220
```

YGKRFVGPITPIIRSLRKELYLVPYHEVDWNKARNQCAKEDLYYPHPLVQDILWATLH
HVYEPLFMHWPAKRLREKALQSVMQHIHYEDENTRYICIGPVNKVLNMLCCWAEDPHS
EAFKLHIPRIYDYLWIAEDGMKMQGYNGSQLWDTAFAVQAIISTELAEEYETTLRKAH
KYIKDSQVLEDCPGDLQSWYRHISKGAWPFSTADHGWPISDCTAEGLKAVLLLSKLPS
EIVGKSIDEQQLYNAVNVILSLQNTDGGFATYELTRSYRWLELMNPAETFGDIVIDYP
YVECSSAAIQALAAFKKLYPGHRRDEIDNCIAEAADFIESIQATDGSWYGSWGVCFTY
GGWFGIRGLVAAGRRYNNCSSLRKACDFLLSKELAAGGWGESYLSCQNKVYTNIKDDR
PHIVNTGWAMLSLIDAGQSERDPTPLHRAARVLINSQMEDGDFPQEEIMGVFNKNCMI
SYSAYRNIFPIWALGEYRSRVLKPLK"

ORIGIN

Alignment Scores:

Pred. No.:	8.92e-282	Length:	2556
Score:	3373.00	Matches:	595
Percent Similarity:	89.12%	Conservative:	85
Best Local Similarity:	77.98%	Mismatches:	75
Query Match:	81.32%	Indels:	8
DB:	8	Gaps:	1

US-09-419-679-6 (1-757) x AB116237 (1-2556)

```
Qy          1 MetTrpLysLeuLysPheAlaGluGlyGlyAsnPro---------------------- 12
Db        123 ATGTGGCAGCTCAAGATTGGGGCAGATACAGTGCCTTCCGATCCTTCCAATGCCGGCGGA 182

Qy         13 TrpLeuArgThrLeuAsnAsnHisValGlyArgGlnValTrpGluPheAspProLysLeu 32
Db        183 TGGCTTTCAACTCTCAACAACCACGTCGGCCGCCAAGTTTGGCATTTTCATCCCGAACTC 242

Qy         33 GlySerProGlnAspLeuLeuGluIleGluLysAlaArgGlnAsnPheHisAspAsnArg 52
Db        243 GGCTCCCCGGAGGATCTTCAGCAAATACAACAGGCTCGTCAGCACTTCTCCGATCACAGA 302

Qy         53 PheThrHisLysHisSerAlaAspLeuLeuMetArgMetGlnPheAlaArgGluAsnPro 72
Db        303 TTCGAGAAGAAGCATAGTGCTGATCTCCTCATGCGAATGCAGTTTGCGAAGGAGAATTCG 362

Qy         73 ThrArgGluValLeuProLysValGlyValLysAspIleGluAspValThrGlnGluIle 92
Db        363 TCCTTTGTGAATCTACCGCAAGTCAAAGTTAAAGATAAAGAAGACGTGACAGAGGAAGCA 422

Qy         93 ValThrLysThrLeuArgArgAlaValSerPheHisSerThrLeuGlnCysHisAspGly 112
Db        423 GTAACGCGCACTCTAAGAAGGGCTATCAATTTTTATTCAACAATCCAGGCAGATGACGGT 482

Qy        113 HisTrpProGlyAspTyrGlyGlyProMetPheLeuMetProGlyLeuValIleThrLeu 132
Db        483 CATTGGCCTGGTGATTATGGCGGCCCAATGTTTCTAATTCCTGGTTTGGTCATTACTCTT 542

Qy        133 SerIleThrGlyAlaLeuAsnThrValLeuThrGluGluHisArgLysGluIleCysArg 152
Db        543 TCCATWACTGGCGCCTTGAATGCGGTCTTATCTACAGAACACCAGCGTGAGATTTGCAGA 602

Qy        153 TyrLeuTyrAsnHisGlnAsnLysAspGlyGlyTrpGlyLeuHisIleGluGlyProSer 172
Db        603 TATCTGTACAATCATCAGAATAAAGATGGGGGTTGGGGTTTACATATTGAGGGCCCAAGC 662

Qy        173 ThrMetPheGlySerValLeuSerTyrIleThrLeuArgLeuGluGlyGluGlyProAsn 192
Db        663 ACAATGTTCGGATCTGTGTTGAATTATGTTACTTTGAGATTGCTTGGTGAAGAAGCTGAA 722

Qy        193 AspGlyGlnGlyGluMetGluLysAlaArgAspTrpIleLeuGlyHisGlyGlyAlaThr 212
Db        723 GATGGGCAGGGTGCTGTGGACAAAGCACGGAAATGGATCTTGGACCATGGTGGTGCAGCT 782

Qy        213 TyrIleThrSerTrpGlyLysMetTrpLeuSerValLeuGlyValTyrGluTrpSerGly 232
Db        783 GCGATTACTTCTTGGGGGAAAATGTGGCTTTCAGTACTTGGAGTATATGAATGGGCTGGA 842

Qy        233 AsnAsnProLeuProProGluIleTrpLeuLeuProTyrMetLeuProPheHisProGly 252
Db        843 AATAATCCACTTCCTCCAGAGCTATGGCTCTTGCCTTATCTTCTCCCTTGTCATCCAGGA 902

Qy        253 ArgMetTrpCysHisCysArgMetValTyrLeuProMetSerTyrLeuTyrGlyLysArg 272
Db        903 AGGATGTGGTGTCACTGTCGAATGGTATACTTGCCTATGTGTTATCTGTATGGCAAGAGA 962

Qy        273 PheValGlyProIleSerProThrValLeuSerLeuArgLysGluLeuTyrThrValPro 292
Db        963 TTTGTTGGTCCTATAACGCCTATAATTAGATCTTTAAGGAAGGAACTTTATCTTGTTCCT 1022

Qy        293 TyrHisAspIleAspTrpAspGlnAlaArgAsnLeuCysAlaLysGluAspLeuTyrTyr 312
Db       1023 TATCATGAAGTTGATTGGAATAAGGCTCGAAATCAATGTGCAAAGGAAGATCTATATTAC 1082

Qy        313 ProHisProLeuValGlnAspIleLeuTrpAlaSerLeuHisLysPheLeuGluProIle 332
Db       1083 CCACATCCGCTGGTTCAGGATATTCTGTGGGCTACGCTGCACCATGTCTATGAGCCTCTT 1142

Qy        333 LeuMetHisTrpProGlyLysArgLeuArgGluLysAlaIleIleSerAlaLeuGluHis 352
Db       1143 TTTATGCATTGGCCTGCAAAAAGATTGAGGGAGAAGGCTTTGCAGAGTGTAATGCAGCAT 1202

Qy        353 IleHisTyrGluAspGluAsnThrArgTyrIleCysIleGlyProValAsnLysValLeu 372
Db       1203 ATCCACTATGAAGATGAAAATACTCGGTATATATGCATAGGGCCTGTCAACAAGGTGTTA 1262

Qy        373 AsnMetLeuCysCysTrpValGluAspProAsnSerGluAlaPheLysLeuHisLeuPro 392
Db       1263 AACATGCTTTGCTGTTGGGCGGAAGATCCACATTCAGAGGCATTCAAATTGCATATTCCT 1322

Qy        393 ArgIleTyrAspTyrLeuTrpIleAlaGluAspGlyMetLysMetGlnGlyTyrAsnGly 412
Db       1323 AGAATCTATGACTATCTGTGGATTGCTGAAGATGGCATGAAAATGCAGGGTTATAATGGA 1382

Qy        413 SerGlnLeuTrpAspThrAlaPheAlaValGlnAlaIleIleAlaSerAsnLeuIleGlu 432
Db       1383 AGTCAATTATGGGATACTGCATTTGCTGTTCAAGCCATCATTTCAACTGAGCTCGCTGAA 1442

Qy        433 GluPheGlyProThrIleArgLysAlaHisThrTyrIleLysAsnSerGlnValLeuGlu 452
Db       1443 GAATATGAGACAACTTTAAGAAAGGCGCACAAGTATATAAAAGACTCCCAGGTCTTAGAA 1502

Qy        453 AspCysProGlyAspLeuAsnLysTrpTyrArgHisIleSerLysGlyAlaTrpProPhe 472
Db       1503 GATTGTCCTGGAGATTTACAAAGTTGGTACCGTCATATTTCAAAAGGTGCATGGCCATTT 1562

Qy        473 SerThrGlyAspHisGlyTrpProIleSerAspCysThrAlaGluGlyLeuLysAlaVal 492
Db       1563 TCAACTGCGGATCACGGATGGCCCATCTCTGACTGCACTGCTGAAGGTTTGAAGGCTGTT 1622

Qy        493 LeuLeuLeuSerLysIleAlaProGluIleValGlyGluProIleAspValLysArgLeu 512
Db       1623 CTTTTATTGTCAAAACTTCCTTCAGAAATAGTAGGGAAGTCAATTGATGAACAACAGCTA 1682

Qy        513 TyrAspSerValAsnValIleLeuSerLeuGlnAsnGluAspGlyGlyPheAlaThrTyr 532
Db       1683 TACAACGCAGTAAATGTCATCCTGTCCTTACAGAATACTGATGGTGGCTTTGCAACTTAT 1742

Qy        533 GluLeuLysArgSerTyrAsnTrpLeuGluIleIleAsnProAlaGluThrPheGlyAsp 552
Db       1743 GAGTTGACCAGGTCATATCGGTGGCTGGAGCTAATGAACCCTGCTGAAACTTTGGCGAT 1802

Qy        553 IleValIleAspTyrProTyrValGluCysThrSerAlaAlaIleGlnAlaLeuAlaSer 572
Db       1803 ATTGTCATTGACTACCCATATGTCGAGTGTTCCTCAGCAGCGATTCAAGCACTTGCAGCA 1862

Qy        573 PheArgLysLeuTyrProGlyHisArgArgGluGluIleGlnHisCysIleAspLysAla 592
Db       1863 TTCAAGAAATTATATCCAGGTCATAGAAGGGATGAAATTGATAATTGTATTGCTGAAGCT 1922

Qy        593 ThrThrPheIleGluLysIleGlnAlaSerAspGlySerTrpTyrGlySerTrpGlyVal 612
Db       1923 GCGGACTTCATTGAAAGTATACAAGCAACTGATGGATCTTGGTATGGATCTTGGGGAGTC 1982

Qy        613 CysPheThrTyrGlyAlaTrpPheGlyValLysGlyLeuIleAlaAlaGlyArgSerPhe 632
```

```
Db        903 TGGCTTCTCCCATATTTCCTGCCAATACATCCAGGAAGGATGTGGTGCCATTGTCGAATG 962

Qy        261 ValTyrLeuProMetSerTyrLeuTyrGlyLysArgPheValGlyProIleSerProThr 280

Db        963 GTGTACTTGCCGATGTCGTATTTGTATGGAAAAAGGTTTGTGGGTCCCATAACGTCCACT 1022

Qy        281 ValLeuSerLeuArgLysGluLeuTyrThrValProTyrHisAspIleAspTrpAspGln 300

Db       1023 GTTTTATCACTGAGAAAGGAGCTTTTCACAGTACCATATCATGAAGTCAACTGGAATGAA 1082

Qy        301 AlaArgAsnLeuCysAlaLysGluAspLeuTyrTyrProHisProLeuValGlnAspIle 320

Db       1083 GCACGCAACCTTTGCGCAAAGGAGGATTTATACTACCCACATCCACTTGTGCAAGATATT 1142

Qy        321 LeuTrpAlaSerLeuHisLysPheLeuGluProIleLeuMetHisTrpProGlyLysArg 340

Db       1143 CTTTGGGCATCACTTCATAAGATTGTTGAGCCTGTTCTGATGCGATGGCCTGGTGCAAAT 1202

Qy        341 LeuArgGluLysAlaIleIleSerAlaLeuGluHisIleHisTyrGluAspGluAsnThr 360

Db       1203 TTGAGAGAAAAGGCTATAAGAACCGCAATAGAACATATTCATTATGAAGATGAGAATACT 1262

Qy        361 ArgTyrIleCysIleGlyProValAsnLysValLeuAsnMetLeuCysCysTrpValGlu 380

Db       1263 AGGTACATCTGCATAGGTCCCGTGAACAAGGTATTAAATATGCTTTGCTGTTGGGTAGAA 1322

Qy        381 AspProAsnSerGluAlaPheLysLeuHisLeuProArgIleTyrAspTyrLeuTrpIle 400

Db       1323 GACCCAAACTCAGAGGCTTTCAAGTTGCACCTACCAAGAATCCATGACTTTCTCTGGTTA 1382

Qy        401 AlaGluAspGlyMetLysMetGlnGlyTyrAsnGlySerGlnLeuTrpAspThrAlaPhe 420

Db       1383 GCTGAAGATGGAATGAAGATGCAGGGTTATAACGGAAGCCAGCTATGGGATACAGGTTTT 1442

Qy        421 AlaValGlnAlaIleIleAlaSerAsnLeuIleGluGluPheGlyProThrIleArgLys 440

Db       1443 GCTATTCAAGCGATTTTGGCAACTAACCTCGTCGAAGAATATGGGCCCGTTTTGGAAAAA 1502

Qy        441 AlaHisThrTyrIleLysAsnSerGlnValLeuGluAspCysProGlyAspLeuAsnLys 460

Db       1503 GCACATTCATTTGTCAAGAATTCCCAGGTGTTAGAAGACTGCCCTGGAGATCTGAATTAC 1562

Qy        461 TrpTyrArgHisIleSerLysGlyAlaTrpProPheSerThrGlyAspHisGlyTrpPro 480

Db       1563 TGGTATCGCCACATTTCTAAAGGGGCTTGGCCTTTCTCAACTGCAGATCACGGTTGGCCC 1622

Qy        481 IleSerAspCysThrAlaGluGlyLeuLysAlaValLeuLeuLeuSerLysIleAlaPro 500

Db       1623 ATCTCTGACTGCACCGCAGAAGGACTGAAAGCTGCTCTTTTGCTATCCAAAGTTCCCAAG 1682

Qy        501 GluIleValGlyGluProIleAspValLysArgLeuTyrAspSerValAsnValIleLeu 520

Db       1683 GCGATTGTTGGTGAACCAATAGATGCAAAACGGTTATATGAAGCTGTTAATGTTATCATT 1742

Qy        521 SerLeuGlnAsnGluAspGlyGlyPheAlaThrTyrGluLeuLysArgSerTyrAsnTrp 540

Db       1743 TCTTTACAGAATGCAGATGGAGGCCTCGCAACATATGAGCTCACCAGGTCATACCCTTGG 1802

Qy        541 LeuGluIleIleAsnProAlaGluThrPheGlyAspIleValIleAspTyrProTyrVal 560

Db       1803 TTAGAGCTAATCAACCCAGCAGAAACCTTTGGCGATATTGTTATTGATTATCCTTACGTG 1862

Qy        561 GluCysThrSerAlaAlaIleGlnAlaLeuAlaSerPheArgLysLeuTyrProGlyHis 580

Db       1863 GAATGTACATCAGCTGCTATCCAAGCTTTGATATCATTTCGAAAGCTGTATCCTGGTCAT 1922

Qy        581 ArgArgGluGluIleGlnHisCysIleAspLysAlaThrThrPheIleGluLysIleGln 600

Db       1923 CGAAAGAAGGAAGTAGATGAGTGCATTGAGAAGGCGGTTAAGTTCATTGAATCCATTCAA 1982

Qy        601 AlaSerAspGlySerTrpTyrGlySerTrpGlyValCysPheThrTyrGlyAlaTrpPhe 620

Db       1983 GCAGCAGATGGCTCATGGTATGGATCATGGGCTGTTTGCTTCACGTATGGTACGTGGTTT 2042

Qy        621 GlyValLysGlyLeuIleAlaAlaGlyArgSerPheSerAsnCysSerSerIleArgLys 640

Db       2043 GGAGTGAAAGGGCTGGTAGCTGTTGGAAAAACATTGAAAAACTCTCCACATGTTGCTAAA 2102

Qy        641 AlaCysGluPheLeuLeuSerLysGlnLeuProSerGlyGlyTrpGlyGluSerTyrLeu 660

Db       2103 GCTTGTGAATTTCTATTGTCGAAACAACAACCTTCGGGCGGCTGGGGAGAAAGCTATCTT 2162

Qy        661 SerCysGlnAsnLysValTyrSerAsnLeuGluGlyAsnArgSerHisValValAsnThr 680

Db       2163 TCATGTCAAGACAAGGTCTATTCAAACCTTGATGGCAACAGATCTCACGTCGTGAATACA 2222

Qy        681 GlyTrpAlaMetLeuAlaLeuIleAspAlaGlyGlnAlaLysArgAspSerGlnProLeu 700

Db       2223 GCATGGGCTATGCTCGCACTCATTGGTGCTGGGCAAGCTGAGGTAGACCGGAAACCACTA 2282

Qy        701 HisArgAlaAlaAlaTyrLeuIleAsnSerGlnLeuGluAspGlyAspProGlnGln 720

Db       2283 CACCGGGCTGCAAGATACTTGATTAATGCTCAAATGGAGAATGGTGATTTTCCACAACAG 2342

Qy        721 GluIleMetGlyValPheAsnLysAsnCysMetIleThrTyrAlaAlaTyrArgAsnIle 740

Db       2343 GAAATAATGGGAGTCTTCAATAGGAACTGCATGATAACATATGCCGCGTATCGAAACATT 2402

Qy        741 ProIleTrpAlaLeuGlyGluTyrGlnSerGlnValLeu 754

Db       2403 TTTCCGATATGGGCTTTGGGGGAGTACCGTTGTCAGGTATTA 2444
```

```
RESULT 11
AB116237
LOCUS       AB116237               2556 bp    mRNA    linear   PLN 03-AUG-2004
DEFINITION  Cucurbita pepo CPX mRNA for cycloartenol synthase, complete cds.
ACCESSION   AB116237
VERSION     AB116237.1  GI:50896400
KEYWORDS    .
SOURCE      Cucurbita pepo
  ORGANISM  Cucurbita pepo
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Cucurbitales; Cucurbitaceae; Cucurbita.
REFERENCE   1
  AUTHORS   Shibuya,M., Adachi,S. and Ebizuka,Y.
  TITLE     Cucurbitadienol Synthase, the First Committed Enzyme for
            Cucurbitacin Biosynthesis, is a Distinct Enzyme from Cycloartenol
            Synthase for Phytosterol Biosynthesis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2556)
  AUTHORS   Shibuya,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-2003) Masaaki Shibuya, The University of Tokyo,
            Graduate School of Pharmaceutical Sciences; Hongo 7-3-1, Bunkyo-ku,
            Tokyo 1130033, Japan (E-mail:shibuyam@mol.f.u-tokyo.ac.jp,
            URL:http://www.f.u-tokyo.ac.jp/~tennen/head.htm,
            Tel:81-3-5841-4741, Fax:81-3-5841-4744)
FEATURES             Location/Qualifiers
     source          1. .2556
                     /organism="Cucurbita pepo"
                     /mol_type="mRNA"
                     /db_xref="taxon:3663"
     gene            1. .2556
                     /gene="CPX"
     CDS             123. .2423
                     /gene="CPX"
                     /codon_start=1
                     /product="cycloartenol synthase"
                     /protein_id="BAD34644.1"
                     /db_xref="GI:50896401"
                     /translation="MWQLKIGADTVPSDPSNAGGWLSTLNNHVGRQVWHFHPELGSPE
                     DLQQIQQARQHFSDHRFEKKHSADLLMRMQFAKENSSFVNLPQVKVKDKEDVTEEAVT
                     RTLRRAINFYSTIQADDGHWPGDYGGPMFLIPGLVITLSITGALNAVLSTEHQREICR
                     YLYNHQNKDGGWGLHIEGPSTMFGSVLNYVTLRLLGEEAEDGQGAVDKARKWILDHGG
                     AAAITSWGKMWLSVLGVYEWAGNNPLPPELWLLPYLLPCHPGRMWCHCRMVYLPMCYL
```

```
RESULT 10
AY094394
LOCUS       AY094394               2654 bp    mRNA    linear   PLN 05-MAY-2002
DEFINITION  Arabidopsis thaliana At2g07050/T4E14.16 mRNA, complete cds.
ACCESSION   AY094394
VERSION     AY094394.1  GI:20453055
KEYWORDS    FLI_CDNA.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 2654)
  AUTHORS   Shinn,P., Chen,H., Cheuk,R., Kim,C.J., Meyers,M.C., Banh,J.,
            Bowser,L., Carninci,P., Chang,E., Dale,J.M., Goldsmith,A.D.,
            Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G.,
            Kawai,J., Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M.,
            Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L., Sakurai,T.,
            Satou,M., Seki,M., Southwick,A., Tang,C.C., Toriumi,M., Wu,H.C.,
            Yamada,K., Yamamura,Y., Yu,G., Yu,S., Shinozaki,K., Davis,R.W.,
            Theologis,A. and Ecker,J.R.
  TITLE     Arabidopsis cDNA clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2654)
  AUTHORS   Shinn,P., Chen,H., Cheuk,R., Kim,C.J., Meyers,M.C., Banh,J.,
            Bowser,L., Carninci,P., Chang,E., Dale,J.M., Goldsmith,A.D.,
            Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G.,
            Kawai,J., Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M.,
            Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L., Sakurai,T.,
            Satou,M., Seki,M., Southwick,A., Tang,C.C., Toriumi,M., Wu,H.C.,
            Yamada,K., Yamamura,Y., Yu,G., Yu,S., Shinozaki,K., Davis,R.W.,
            Theologis,A. and Ecker,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-APR-2002) Salk Institute Genomic Analysis Laboratory
            (SIGnAL), Plant Biology Laboratory, The Salk Institute for
            Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
            USA
COMMENT     RIKEN Genomic Sciences Center (GSC) members carried out the
            collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
            Arabidopsis Full-Length cDNA') :  Seki,M., Narusaka,M., Ishida,J.,
            Satou,M., Kamiya,A., Sakurai,T., Carninci,P.,  Kawai,J.,
            Hayashizaki,Y. and Shinozaki,K.

            The Salk, Stanford, PGEC (SSP) Consortium members carried out the
            sequencing and annotation of the RAFL cDNAs: Shinn,P., Chen,H.,
            Cheuk,R., Kim,C.J.,  Meyers,M.C., Banh,J. Bowser,L., Chang,E.,
            Dale,J.M., Goldsmith,A.D., Jones,T., Karlin-Neumann,G., Lam,B.,
            Lee,J.M., Lin,J., Miranda,M., Nguyen,M., Onodera,C.S., Palm,C.J.,
            Quach,H.L., Southwick,A., Tang,C.C., Toriumi,M., Wu,H.C.,
            Yamada,K., Yamamura, Y., Yu,G., Yu,S., Davis,R.W.,  Theologis,A.,
            and Ecker,J.R.

            Shinn,P. (SSP/Salk) and Seki,M. (RIKEN GSC) contributed equally to
            this work.  Shinozaki,K. (RIKEN GSC) and Ecker,J.R. (SSP/Salk)
            contributed equally to this work as PIs.
FEATURES             Location/Qualifiers
     source          1. .2654
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /chromosome="2"
                     /clone="RAFL09-12-N05(R21897)"
                     /ecotype="Columbia"
     CDS             183. .2462
                     /note="cycloartenol synthase"
                     /codon_start=1
                     /product="At2g07050/T4E14.16"
                     /protein_id="AAM19773.1"
                     /db_xref="GI:20453056"
                     /translation="MWKLKIAEGGSPWLRTTNNHVGRQFWEFDPNLGTPEDLAAVEEA
                     RKSFSDNRFVQKHSADLLMRLQFSRENLISPVLPQVKIEDTDDVTEEMVETTLKRGLD
                     FYSTIQAHDGHWPGDYGGPMFLLPGLIITLSITGALNTVLSEQHKQEMRRYLYNHQNE
                     DGGWGLHIEGPSTMFGSVLNYVTLRLLGEGPNDGDGDMEKGRDWILNHGGATNITSWG
                     KMWLSVLGAFEWSGNNPLPPEIWLLPYFLPIHPGRMWCHCRMVYLPMSYLYGKRFVGP
                     ITSTVLSLRKELFTVPYHEVNWNEARNLCAKEDLYYPHPLVQDILWASLHKIVEPVLM
                     RWPGANLREKAIRTAIEHIHYEDENTRYICIGPVNKVLNMLCCWVEDPNSEAFKLHLP
                     RIHDFLWLAEDGMKMQGYNGSQLWDTGFAIQAILATNLVEEYGPVLEKAHSFVKNSQV
                     LEDCPGDLNYWYRHISKGAWPFSTADHGWPISDCTAEGLKAALLLSKVPKAIVGEPID
                     AKRLYEAVNVIISLQNADGGLATYELTRSYPWLELINPAETFGDIVIDYPYVECTSAA
                     IQALISFRKLYPGHRKKEVDECIEKAVKFIESIQAADGSWYGSWAVCFTYGTWFGVKG
                     LVAVGKTLKNSPHVAKACEFLLSKQQPSGGWGESYLSCQDKVYSNLDGNRSHVVNTAW
                     AMLALIGAGQAEVDRKPLHRAARYLINAQMENGDFPQQEIMGVFNRNCMITYAAYRNI
                     FPIWALGEYRCQVLLQQGE"
ORIGIN

Alignment Scores:
Pred. No.:                5.13e-291       Length:          2654
Score:                    3480.00         Matches:         607
Percent Similarity:       90.72%          Conservative:    77
Best Local Similarity:    80.50%          Mismatches:      70
Query Match:              83.90%          Indels:          0
DB:                       8               Gaps:            0

US-09-419-679-6 (1-757) x AY094394 (1-2654)

Qy          1 MetTrpLysLeuLysPheAlaGluGlyGlyAsnProTrpLeuArgThrLeuAsnAsnHis 20
              ||||||||||||||||   ||||||||||||:::|||||||||||||||||   |||||||||
Db        183 ATGTGGAAACTGAAGATCGCGGAAGGAGGTAGTCCATGGCTTAGAACCACCAATAATCAC 242

Qy         21 ValGlyArgGlnValTrpGluPheAspProLysLeuGlySerProGlnAspLeuLeuGlu 40
              ||||||||||||   ||||||||||||||||   ||||||:::|||:::|||||||
Db        243 GTCGGAAGACAGTTTTGGGAGTTCGATCCGAATCTCGGTACTCCTGAGGATCTCGCCGCC 302

Qy         41 IleGluLysAlaArgGlnAsnPheHisAspAsnArgPheThrHisLysHisSerAlaAsp 60
              :::|||:::||||||:::::|||   ||||||||||||   ||||||||||||||||
Db        303 GTCGAAGAAGCTAGGAAGTCTTTTTCGGATAATCGATTCGTGCAGAAACATAGCGCCGAT 362

Qy         61 LeuLeuMetArgMetGlnPheAlaArgGluAsnProThrArgGluValLeuProLysVal 80
              ||||||||||||:::||||||:::||||||||   ||||||||||:::|||
Db        363 CTGCTTATGCGCCTTCAGTTTTCAAGAGAAAATTTGATTAGCCCAGTTTTACCTCAAGTC 422

Qy         81 GlyValLysAspIleGluAspValThrGlnGluIleValThrLysThrLeuArgArgAla 100
              :::::::|||   :::||||||||:::|||:::|||      |||||||:::|||
Db        423 AAAATCGAAGACACTGATGATGTTACAGAGGAGATGGTGGAAACCACGTTAAAGAGGGGT 482

Qy        101 ValSerPheHisSerThrLeuGlnCysHisAspGlyHisTrpProGlyAspTyrGlyGly 120
              :::   |||:::||||||:::|||   ||||||||||||||||||||||||||||||||
Db        483 CTAGATTTCTATTCAACTATACAGGCACACGACGGGCACTGGCCAGGTGATTATGGTGGT 542

Qy        121 ProMetPheLeuMetProGlyLeuValIleThrLeuSerIleThrGlyAlaLeuAsnThr 140
              ||||||||||||:::||||||||:::|||||||||||||||||||||||||||||||||
Db        543 CCTATGTTTCTTCTCCCAGGACTGATAATTACACTCTCCATAACTGGAGCACTGAATACA 602

Qy        141 ValLeuThrGluGluHisArgLysGluIleCysArgTyrLeuTyrAsnHisGlnAsnLys 160
              ||||||:::|||:::|||:::::::|||:::   |||||||||||||||||||||||:::
Db        603 GTATTGTCGGAACAACATAAACAAGAAATGCGCCGTTATCTCTATAATCACCAGAATGAG 662

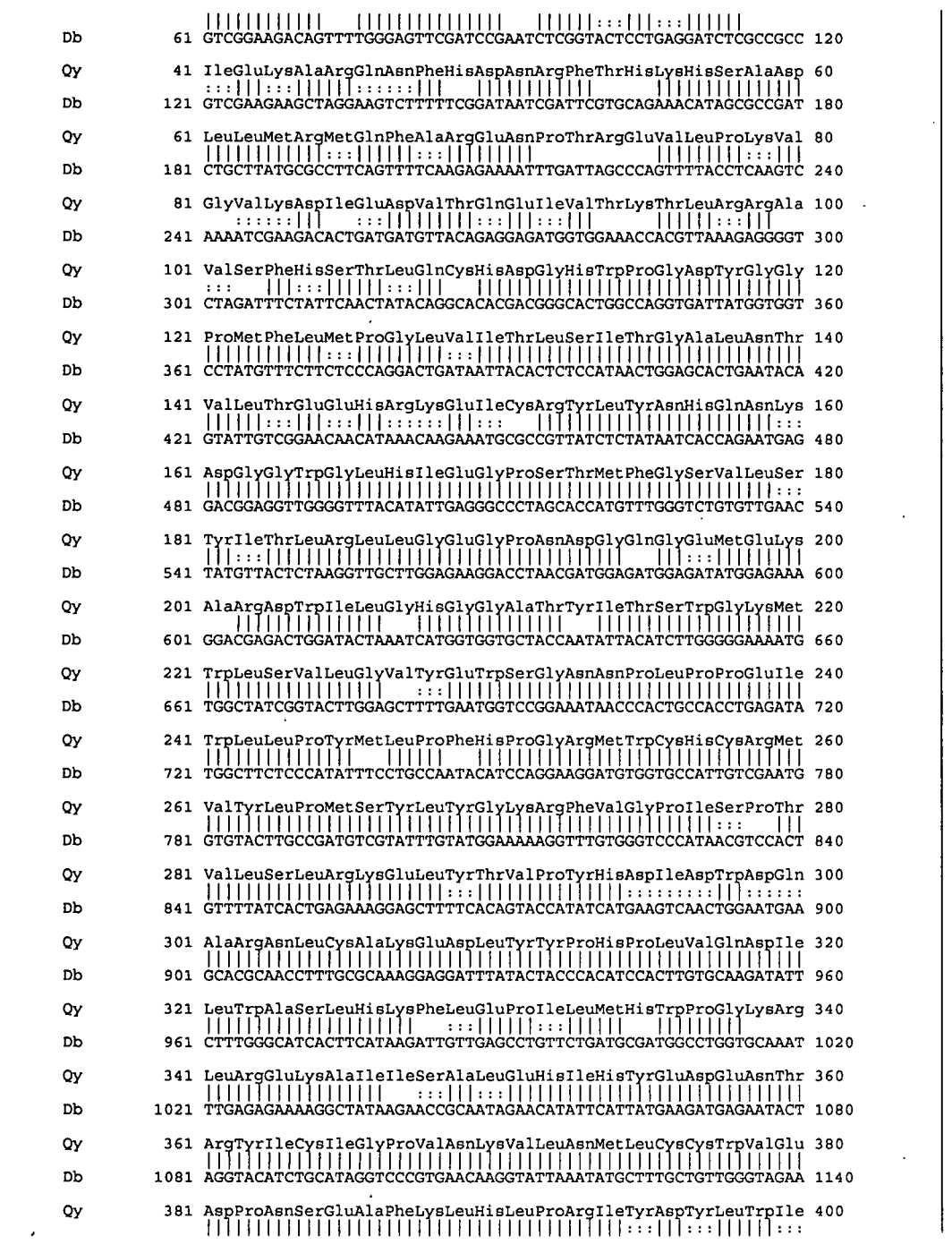
Qy        161 AspGlyGlyTrpGlyLeuHisIleGluGlyProSerThrMetPheGlySerValLeuSer 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        663 GACGGAGGTTGGGGTTTACATATTGAGGGCCCTAGCACCATGTTTGGGTCTGTGTTGAAC 722

Qy        181 TyrIleThrLeuArgLeuLeuGlyGluGlyProAsnAspGlyGlnGlyGluMetGluLys 200
              |||:::|||||||||||||||||||||||||||||||||||   ||:::|||||||||
Db        723 TATGTTACTCTAAGGTTGCTTGGAGAAGGACCTAACGATGGAGATGGAGATATGGAGAAA 782

Qy        201 AlaArgAspTrpIleLeuGlyHisGlyGlyAlaThrTyrIleThrSerTrpGlyLysMet 220
              ||||||||||||||||   ||||||||||||||||   ||||||||||||||||||||
Db        783 GGACGAGACTGGATACTAAATCATGGTGGTGCTACCAATATTACATCTTGGGGGAAAATG 842

Qy        221 TrpLeuSerValLeuGlyValTyrGluTrpSerGlyAsnAsnProLeuProProGluIle 240
              ||||||||||||||||||   :::||||||||||||||||||||||||||||||||||||
Db        843 TGGCTATCGGTACTTGGAGCTTTTGAATGGTCCGGAAATAACCCACTGCCACCTGAGATA 902

Qy        241 TrpLeuLeuProTyrMetLeuProPheHisProGlyArgMetTrpCysHisCysArgMet 260
              ||||||||||||||||   ||||||   ||||||||||||||||||||||||||||||||
```


Db 61 GTCGGAAGACAGTTTTGGGAGTTCGATCCGAATCTCGGTACTCCTGAGGATCTCGCCGCC 120
Qy 41 IleGluLysAlaArgGlnAsnPheHisAspAsnArgPheThrHisLysHisSerAlaAsp 60
Db 121 GTCGAAGAAGCTAGGAAGTCTTTTTCGGATAATCGATTCGTGCAGAAACATAGCGCCGAT 180
Qy 61 LeuLeuMetArgMetGlnPheAlaArgGluAsnProThrArgGluValLeuProLysVal 80
Db 181 CTGCTTATGCGCCTTCAGTTTTCAAGAGAAAATTTGATTAGCCCAGTTTTACCTCAAGTC 240
Qy 81 GlyValLysAspIleGluAspValThrGlnGluIleValThrLysThrLeuArgArgAla 100
Db 241 AAAATCGAAGACACTGATGATGTTACAGAGGAGATGGTGGAAACCACGTTAAAGAGGGGT 300
Qy 101 ValSerPheHisSerThrLeuGlnCysHisAspGlyHisTrpProGlyAspTyrGlyGly 120
Db 301 CTAGATTTCTATTCAACTATACAGGCACACGACGGGCACTGGCCAGGTGATTATGGTGGT 360
Qy 121 ProMetPheLeuMetProGlyLeuValIleThrLeuSerIleThrGlyAlaLeuAsnThr 140
Db 361 CCTATGTTTCTTCTCCCAGGACTGATAATTACACTCTCCATAACTGGAGCACTGAATACA 420
Qy 141 ValLeuThrGluGluHisArgLysGluIleCysArgTyrLeuTyrAsnHisGlnAsnLys 160
Db 421 GTATTGTCGGAACAACATAAACAAGAAATGCGCCGTTATCTCTATAATCACCAGAATGAG 480
Qy 161 AspGlyGlyTrpGlyLeuHisIleGluGlyProSerThrMetPheGlySerValLeuSer 180
Db 481 GACGGAGGTTGGGGTTTACATATTGAGGGCCCTAGCACCATGTTTGGGTCTGTGTTGAAC 540
Qy 181 TyrIleThrLeuArgLeuLeuGlyGluGlyProAsnAspGlyGlnGlyGluMetGluLys 200
Db 541 TATGTTACTCTAAGGTTGCTTGGAGAAGGACCTAACGATGGAGATGGAGATATGGAGAAA 600
Qy 201 AlaArgAspTrpIleLeuGlyHisGlyGlyAlaThrTyrIleThrSerTrpGlyLysMet 220
Db 601 GGACGAGACTGGATACTAAATCATGGTGGTGCTACCAATATTACATCTTGGGGGAAAATG 660
Qy 221 TrpLeuSerValLeuGlyValTyrGluTrpSerGlyAsnAsnProLeuProProGluIle 240
Db 661 TGGCTATCGGTACTTGGAGCTTTTGAATGGTCCGGAAATAACCCACTGCCACCTGAGATA 720
Qy 241 TrpLeuLeuProTyrMetLeuProPheHisProGlyArgMetTrpCysHisCysArgMet 260
Db 721 TGGCTTCTCCCATATTTCCTGCCAATACATCCAGGAAGGATGTGGTGCCATTGTCGAATG 780
Qy 261 ValTyrLeuProMetSerTyrLeuTyrGlyLysArgPheValGlyProIleSerProThr 280
Db 781 GTGTACTTGCCGATGTCGTATTTGTATGGAAAAAGGTTTGTGGGTCCCATAACGTCCACT 840
Qy 281 ValLeuSerLeuArgLysGluLeuTyrThrValProTyrHisAspIleAspTrpAspGln 300
Db 841 GTTTTATCACTGAGAAAGGAGCTTTTCACAGTACCATATCATGAAGTCAACTGGAATGAA 900
Qy 301 AlaArgAsnLeuCysAlaLysGluAspLeuTyrTyrProHisProLeuValGlnAspIle 320
Db 901 GCACGCAACCTTTGCGCAAAGGAGGATTTATACTACCCACATCCACTTGTGCAAGATATT 960
Qy 321 LeuTrpAlaSerLeuHisLysPheLeuGluProIleLeuMetHisTrpProGlyLysArg 340
Db 961 CTTTGGGCATCACTTCATAAGATTGTTGAGCCTGTTCTGATGCGATGGCCTGGTGCAAAT 1020
Qy 341 LeuArgGluLysAlaIleIleSerAlaLeuGluHisIleHisTyrGluAspGluAsnThr 360
Db 1021 TTGAGAGAAAAGGCTATAAGAACCGCAATAGAACATATTCATTATGAAGATGAGAATACT 1080
Qy 361 ArgTyrIleCysIleGlyProValAsnLysValLeuAsnMetLeuCysCysTrpValGlu 380
Db 1081 AGGTACATCTGCATAGGTCCCGTGAACAAGGTATTAAATATGCTTTGCTGTTGGGTAGAA 1140
Qy 381 AspProAsnSerGluAlaPheLysLeuHisLeuProArgIleTyrAspTyrLeuTrpIle 400
Db 1141 GACCCAAACTCAGAGGCTTTCAAGTTGCACCTACCAAGAATCCATGACTTTCTCTGGTTA 1200
Qy 401 AlaGluAspGlyMetLysMetGlnGlyTyrAsnGlySerGlnLeuTrpAspThrAlaPhe 420
Db 1201 GCTGAAGATGGAATGAAGATGCAGGGTTATAACGGAAGCCAGCTATGGGATACAGGTTTT 1260
Qy 421 AlaValGlnAlaIleIleAlaSerAsnLeuIleGluGluPheGlyProThrIleArgLys 440
Db 1261 GCTATTCAAGCGATTTTGGCAACTAACCTCGTCGAAGAATATGGGCCCGTTTTGGAAAAA 1320
Qy 441 AlaHisThrTyrIleLysAsnSerGlnValLeuGluAspCysProGlyAspLeuAsnLys 460
Db 1321 GCACATTCATTTGTCAAGAATTCCCAGGTGTTAGAAGACTGCCCTGGAGATCTGAATTAC 1380
Qy 461 TrpTyrArgHisIleSerLysGlyAlaTrpProPheSerThrGlyAspHisGlyTrpPro 480
Db 1381 TGGTATCGCCACATTTCTAAAGGGGCTTGGCCTTTCTCAACTGCAGATCACGGTTGGCCC 1440
Qy 481 IleSerAspCysThrAlaGluGlyLeuLysAlaValLeuLeuLeuSerLysIleAlaPro 500
Db 1441 ATCTCTGACTGCACCGCAGAAGGACTGAAAGCTGCTCTTTTGCTATCCAAAGTTCCCAAG 1500
Qy 501 GluIleValGlyGluProIleAspValLysArgLeuTyrAspSerValAsnValIleLeu 520
Db 1501 GCGATTGTTGGTGAACCAATAGATGCAAAACGGTTATATGAAGCTGTTAATGTTATCATT 1560
Qy 521 SerLeuGlnAsnGluAspGlyGlyPheAlaThrTyrGluLeuLysArgSerTyrAsnTrp 540
Db 1561 TCTTTACAGAATGCAGATGGAGGCCTCGCAACATATGAGCTCACCAGGTCATACCCTTGG 1620
Qy 541 LeuGluIleIleAsnProAlaGluThrPheGlyAspIleValIleAspTyrProTyrVal 560
Db 1621 TTAGAGCTAATCAACCCAGCAGAAACCTTTGGCGATATTGTTATTGATTATCCTTACGTG 1680
Qy 561 GluCysThrSerAlaAlaIleGlnAlaLeuAlaSerPheArgLysLeuTyrProGlyHis 580
Db 1681 GAATGTACATCAGCTGCTATCCAAGCTTTGATATCATTTCGAAAGCTGTATCCTGGTCAT 1740
Qy 581 ArgArgGluGluIleGlnHisCysIleAspLysAlaThrThrPheIleGluLysIleGln 600
Db 1741 CGAAAGAAGGAAGTAGATGAGTGCATTGAGAAGGCGGTTAAGTTCATTGAATCCATTCAA 1800
Qy 601 AlaSerAspGlySerTrpTyrGlySerTrpGlyValCysPheThrTyrGlyAlaTrpPhe 620
Db 1801 GCAGCAGATGGCTCATGGTATGGATCATGGGCTGTTTGCTTCACGTATGGTACGTGGTTT 1860
Qy 621 GlyValLysGlyLeuIleAlaAlaGlyArgSerPheSerAsnCysSerSerIleArgLys 640
Db 1861 GGAGTGAAAGGGCTGGTAGCTGTTGGAAAAACATTGAAAAACTCTCCACATGTTGCTAAA 1920
Qy 641 AlaCysGluPheLeuLeuSerLysGlnLeuProSerGlyGlyTrpGlyGluSerTyrLeu 660
Db 1921 GCTTGTGAATTTCTATTGTCGAAACAACAACCTTCGGGCGGCTGGGGAGAAAGCTATCTT 1980
Qy 661 SerCysGlnAsnLysValTyrSerAsnLeuGluGlyAsnArgSerHisValValAsnThr 680
Db 1981 TCATGTCAAGACAAGGTCTATTCAAACCTTGATGGCAACAGATCTCACGTCGTGAATACA 2040
Qy 681 GlyTrpAlaMetLeuAlaLeuIleAspAlaGlyGlnAlaLysArgAspSerGlnProLeu 700
Db 2041 GCATGGGCTATGCTCGCACTCATTGGTGCTGGGCAAGCTGAGGTAGACCGGAAACCACTA 2100
Qy 701 HisArgAlaAlaAlaTyrLeuIleAsnSerGlnLeuGluAspGlyAspProProGlnGln 720
Db 2101 CACCGGGCTGCAAGATACTTGATTAATGCTCAAATGGAGAATGGTGATTTTCCACAACAG 2160
Qy 721 GluIleMetGlyValPheAsnLysAsnCysMetIleThrTyrAlaAlaTyrArgAsnIle 740
Db 2161 GAAATAATGGGAGTCTTCAATAGGAACTGCATGATAACATATGCCGCGTATCGAAACATT 2220
Qy 741 ProIleTrpAlaLeuGlyGluTyrGlnSerGlnValLeu 754
Db 2221 TTTCCGATATGGGCTTTGGGGGAGTACCGTTGTCAGGTATTA 2262

```
Qy        441 AlaHisThrTyrIleLysAsnSerGlnValLeuGluAspCysProGlyAspLeuAsnLys 460
              ||||||:::::::::|||||||||||||||||||||||||||||||||||||||||||||
Db       1321 GCACATTCATTTGTCAAGAATTCCCAGGTGTTAGAAGACTGCCCTGGAGATCTGAATTAC 1380

Qy        461 TrpTyrArgHisIleSerLysGlyAlaTrpProPheSerThrGlyAspHisGlyTrpPro 480
              ||||||||||||||||||||||||||||||||||||||||   ||||||||||||||||
Db       1381 TGGTATCGCCACATTTCTAAAGGGGCTTGGCCTTTCTCAACTGCAGATCACGGTTGGCCC 1440

Qy        481 IleSerAspCysThrAlaGluGlyLeuLysAlaValLeuLeuLeuSerLysIleAlaPro 500
              |||||||||||||||||||||||||||||||   ||||||||||||||:::
Db       1441 ATCTCTGACTGCACCGCAGAAGGACTGAAAGCTGCTCTTTTGCTATCCAAAGTTCCCAAG 1500

Qy        501 GluIleValGlyGluProIleAspValLysArgLeuTyrAspSerValAsnValIleLeu 520
                 ||||||||||||||||||||   ||||||||||||:::::|||||||||||||:::
Db       1501 GCGATTGTTGGTGAACCAATAGATGCAAAACGGTTATATGAAGCTGTTAATGTTATCATT 1560

Qy        521 SerLeuGlnAsnGluAspGlyGlyPheAlaThrTyrGluLeuLysArgSerTyrAsnTrp 540
              ||||||||||||   |||||||||   ||||||||||||||   |||||||||   |||
Db       1561 TCTTTACAGAATGCAGATGGAGGCCTCGCAACATATGAGCTCACCAGGTCATACCCTTGG 1620

Qy        541 LeuGluIleIleAsnProAlaGluThrPheGlyAspIleValIleAspTyrProTyrVal 560
              ||||||:::||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TTAGAGCTAATCAACCCAGCAGAAACCTTTGGCGATATTGTTATTGATTATCCTTACGTG 1680

Qy        561 GluCysThrSerAlaAlaIleGlnAlaLeuAlaSerPheArgLysLeuTyrProGlyHis 580
              ||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||
Db       1681 GAATGTACATCAGCTGCTATCCAAGCTTTGATATCATTTCGAAAGCTGTATCCTGGTCAT 1740

Qy        581 ArgArgGluGluIleGlnHisCysIleAspLysAlaThrThrPheIleGluLysIleGln 600
              |||::::::|||:::      |||||||:::||||||      |||||||||   ||||||
Db       1741 CGAAAGAAGGAAGTAGATGAGTGCATTGAGAAGGCGGTTAAGTTCATTGAATCCATTCAA 1800

Qy        601 AlaSerAspGlySerTrpTyrGlySerTrpGlyValCysPheThrTyrGlyAlaTrpPhe 620
              |||:::|||||||||||||||||||||||||   ||||||||||||||||||   ||||||
Db       1801 GCAGCAGATGGCTCATGGTATGGATCATGGGCTGTTTGCTTCACGTATGGTACGTGGTTT 1860

Qy        621 GlyValLysGlyLeuIleAlaAlaGlyArgSerPheSerAsnCysSerSerIleArgLys 640
              ||||||||||||||:::|||   |||::::::      |||         :::   |||
Db       1861 GGAGTGAAAGGGCTGGTAGCTGTTGGAAAAACATTGAAAAACTCTCCACATGTTGCTAAA 1920

Qy        641 AlaCysGluPheLeuLeuSerLysGlnLeuProSerGlyGlyTrpGlyGluSerTyrLeu 660
              |||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||
Db       1921 GCTTGTGAATTTCTATTGTCGAAACAACAACCTTCGGGCGGCTGGGGAGAAAGCTATCTT 1980

Qy        661 SerCysGlnAsnLysValTyrSerAsnLeuGluGlyAsnArgSerHisValValAsnThr 680
              |||||||||:::|||||||||||||||||||:::|||||||||||||||||||||||||
Db       1981 TCATGTCAAGACAAGGTCTATTCAAACCTTGATGGCAACAGATCTCACGTCGTGAATACA 2040

Qy        681 GlyTrpAlaMetLeuAlaLeuIleAspAlaGlyGlnAlaLysArgAspSerGlnProLeu 700
                 ||||||||||||||||||||   ||||||||||||:::   |||   :::||||||
Db       2041 GCATGGGCTATGCTCGCACTCATTGGTGCTGGGCAAGCTGAGGTAGACCGGAAACCACTA 2100

Qy        701 HisArgAlaAlaAlaTyrLeuIleAsnSerGlnLeuGluAspGlyAspPheProGlnGln 720
              ||||||||||||   |||||||||||:::|||:::|||:::||||||||||||||||||
Db       2101 CACCGGGCTGCAAGATACTTGATTAATGCTCAAATGGAGAATGGTGATTTTCCACAACAG 2160

Qy        721 GluIleMetGlyValPheAsnLysAsnCysMetIleThrTyrAlaAlaTyrArgAsnIle 740
              ||||||||||||||||||:::||||||||||||||||||||||||||||||||||||||
Db       2161 GAAATAATGGGAGTCTTCAATAGGAACTGCATGATAACATATGCCGCGTATCGAAACATT 2220

Qy        741 PheProIleTrpAlaLeuGlyGluTyrGlnSerGlnValLeu 754
              |||||||||||||||||||||||||:::   |||||||||
Db       2221 TTTCCGATATGGGCTTTGGGGGAGTACCGTTGTCAGGTATTA 2262

RESULT 9
BT001118
LOCUS       BT001118               2280 bp    mRNA    linear   PLN 08-NOV-2002
DEFINITION  Arabidopsis thaliana At2g07050/T4E14.16 mRNA, complete cds.
ACCESSION   BT001118
VERSION     BT001118.1  GI:24796993
KEYWORDS    FLI_CDNA.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 2280)
  AUTHORS   Shinn,P., Chen,H., Cheuk,R., Kim,C.J., Bowser,L., Carninci,P.,
            Chan,M.M., Chang,C.H., Dale,J.M., Hayashizaki,Y., Hsuan,V.W.,
            Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J.,
            Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M., Nguyen,M.,
            Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M., Southwick,A.,
            Tang,C.C., Toriumi,M., Wallender,E.K., Wong,C., Wu,H.C., Yamada,K.,
            Yu,G., Yuan,S., Shinozaki,K., Davis,R.W., Theologis,A. and
            Ecker,J.R.
  TITLE     Arabidopsis ORF clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2280)
  AUTHORS   Shinn,P., Chen,H., Cheuk,R., Kim,C.J., Bowser,L., Carninci,P.,
            Chan,M.M., Chang,C.H., Dale,J.M., Hayashizaki,Y., Hsuan,V.W.,
            Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J.,
            Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M., Nguyen,M.,
            Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M., Southwick,A.,
            Tang,C.C., Toriumi,M., Wallender,E.K., Wong,C., Wu,H.C., Yamada,K.,
            Yu,G., Yuan,S., Shinozaki,K., Davis,R.W., Theologis,A. and
            Ecker,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-2002) Salk Institute Genomic Analysis Laboratory
            (SIGnAL), Plant Biology Laboratory, The Salk Institute for
            Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
            USA
FEATURES             Location/Qualifiers
     source          1. .2280
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /chromosome="2"
                     /clone="U21897"
                     /ecotype="Columbia"
     CDS             1. .2280
                     /note="cycloartenol synthase"
                     /codon_start=1
                     /product="At2g07050/T4E14.16"
                     /protein_id="AAN64509.1"
                     /db_xref="GI:24796994"
                     /translation="MWKLKIAEGGSPWLRTTNNHVGRQFWEFDPNLGTPEDLAAVEEA
                     RKSFSDNRFVQKHSADLLMRLQFSRENLISPVLPQVKIEDTDDVTEEMVETTLKRGLD
                     FYSTIQAHDGHWPGDYGGPMFLLPGLIITLSITGALNTVLSEQHKQEMRRYLYNHQNE
                     DGGWGLHIEGPSTMFGSVLNYVTLRLLGEGPNDGDGDMEKGRDWILNHGGATNITSWG
                     KMWLSVLGAFEWSGNNPLPPEIWLLPYFLPIHPGRMWCHCRMVYLPMSYLYGKRFVGP
                     ITSTVLSLRKELFTVPYHEVNWNSARNLCAKEDLYYPHPLVQDILWASLHKIVEPVLM
                     RWPGANLREKAIRTAIEHIHYEDENTRYICIGPVNKVLNMLCCWVEDPNSEAFKLHLP
                     RIHDFLWLAEDGMKMQGYNGSQLWDTGFAIQAILATNLVEEYGPVLEKAHSFVKNSQV
                     LEDCPGDLNYWYRHISKGAWPFSTADHGWPISDCTAEGLKAALLLSKVPKAIVGEPID
                     AKRLYEAVNVIISLQNADGGLATYELTRSYPWLELINPAETFGDIVIDYPYVECTSAA
                     IQALISFRKLYPGHRKKEVDECIEKAVKFIESIQAADGSWYGSWAVCFTYGTWFGVKG
                     LVAVGKTLKNSPHVAKACEFLLSKQQPSGGWGESYLSCQDKVYSNLDGNRSHVVNTAW
                     AMLALIGAGQAEVDRKPLHRAARYLINAQMENGDFPQQEIMGVFNRNCMITYAAYRNI
                     FPIWALGEYRCQVLLQQGE"
ORIGIN

Alignment Scores:
Pred. No.:                4.2e-291        Length:          2280
Score:                    3480.00         Matches:         607
Percent Similarity:       90.72%          Conservative:    77
Best Local Similarity:    80.50%          Mismatches:      70
Query Match:              83.90%          Indels:          0
DB:                       8               Gaps:            0

US-09-419-679-6 (1-757) x BT001118 (1-2280)

Qy          1 MetTrpLysLeuLysPheAlaGluGlyGlyAsnProTrpLeuArgThrLeuAsnAsnHis 20
              ||||||||||||||   ||||||||||||:::|||||||||||||||   |||||||||
Db          1 ATGTGGAAACTGAAGATCGCGGAAGGAGGTAGTCCATGGCTTAGAACCACCAATAATCAC 60

Qy         21 ValGlyArgGlnValTrpGluPheAspProLysLeuGlySerProGlnAspLeuLeuGlu 40
```

```
Db       1974 GCTTGTGAATTTCTATTGTCGAAACAACAACCTTCGGGCGGCTGGGGAGAAAGCTATCTT 2033

Qy        661 SerCysGlnAsnLysValTyrSerAsnLeuGluGlyAsnArgSerHisValValAsnThr 680
              |||||||||:::|||||||||||||||||||:::|||||||||||||||||||||||||||
Db       2034 TCATGTCAAGACAAGGTCTATTCAAACCTTGATGGCAACAGATCTCACGTCGTGAATACA 2093

Qy        681 GlyTrpAlaMetLeuAlaLeuIleAspAlaGlyGlnAlaLysArgAspSerGlnProLeu 700
              ||||||||||||||||||||    ||||||||||||:::   |||   :::|||||||
Db       2094 GCATGGGCTATGCTCGCACTCATTGGTGCTGGGCAAGCTGAGGTAGACCGGAAACCACTA 2153

Qy        701 HisArgAlaAlaAlaTyrLeuIleAsnSerGlnLeuGluAspGlyAspPheProGlnGln 720
              ||||||||||||    ||||||||||||:::|||:::|||:::||||||||||||||||
Db       2154 CACCGGGCTGCAAGATACTTGATTAATGCTCAAATGGAGAATGGTGATTTTCCACAACAG 2213

Qy        721 GluIleMetGlyValPheAsnLysAsnCysMetIleThrTyrAlaAlaTyrArgAsnIle 740
              ||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db       2214 GAAATAATGGGAGTCTTCAATAGGAACTGCATGATAACATATGCCGCGTATCGAAACATT 2273

Qy        741 PheProIleTrpAlaLeuGlyGluTyrGlnSerGlnValLeu 754
              |||||||||||||||||||||||||:::   |||||||||
Db       2274 TTTCCGATATGGGCGTTGGGGGAGTACCGTTGTCAGGTATTA 2315

RESULT 8
AX506138
LOCUS       AX506138               2280 bp    DNA     linear   PAT 27-SEP-2002
DEFINITION  Sequence 833 from Patent WO0216655.
ACCESSION   AX506138
VERSION     AX506138.1  GI:23387375
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Harper,J.F., Kreps,J., Wang,X. and Zhu,T.
  TITLE     Stress-regulated genes of plants, transgenic plants containing
            same, and methods of use
  JOURNAL   Patent: WO 0216655-A 833 28-FEB-2002;
            The Scripps Research Institute (US) ; Syngenta Participations AG
            (CH)
FEATURES             Location/Qualifiers
     source          1. .2280
                     /organism="Arabidopsis thaliana"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:3702"
ORIGIN

Alignment Scores:
Pred. No.:              4.2e-291        Length:         2280
Score:                  3480.00         Matches:        607
Percent Similarity:     90.72%          Conservative:   77
Best Local Similarity:  80.50%          Mismatches:     70
Query Match:            83.90%          Indels:         0
DB:                     6               Gaps:           0

US-09-419-679-6 (1-757) x AX506138 (1-2280)

Qy          1 MetTrpLysLeuLysPheAlaGluGlyGlyAsnProTrpLeuArgThrLeuAsnAsnHis 20
              |||||||||||||||   ||||||||||:::||||||||||||||   |||||||||
Db          1 ATGTGGAAACTGAAGATCGCGGAAGGAGGTAGTCCATGGCTTAGAACCACCAATAATCAC 60

Qy         21 ValGlyArgGlnValTrpGluPheAspProLysLeuGlySerProGlnAspLeuLeuGlu 40
              ||||||||||||   ||||||||||||||   ||||||:::|||:::|||||||
Db         61 GTCGGAAGACAGTTTTGGGAGTTCGATCCGAATCTCGGTACTCCTGAGGATCTCGCCGCC 120

Qy         41 IleGluLysAlaArgGlnAsnPheHisAspAsnArgPheThrHisLysHisSerAlaAsp 60
              :::|||:::||||||::::::|||   ||||||||||||      ||||||||||||||||
Db        121 GTCGAAGAAGCTAGGAAGTCTTTTTCGGATAATCGATTCGTGCAGAAACATAGCGCCGAT 180

Qy         61 LeuLeuMetArgMetGlnPheAlaArgGluAsnProThrArgGluValLeuProLysVal 80
              ||||||||||||:::|||||||:::|||||||||          |||||||||:::|||
Db        181 CTGCTTATGCGCCTTCAGTTTTCAAGAGAAAATTTGATTAGCCCAGTTTTACCTCAAGTC 240

Qy         81 GlyValLysAspIleGluAspValThrGlnGluIleValThrLysThrLeuArgArgAla 100
              ::::::|||   :::|||||||||:::|||:::|||      ||||||:::|||
Db        241 AAAATCGAAGACACTGATGATGTTACAGAGGAGATGGTGGAAACCACGTTAAAGAGGGGT 300

Qy        101 ValSerPheHisSerThrLeuGlnCysHisAspGlyHisTrpProGlyAspTyrGlyGly 120
              :::   |||:::|||||||:::|||   ||||||||||||||||||||||||||||||
Db        301 CTAGATTTCTATTCAACTATACAGGCACACGACGGGCACTGGCCAGGTGATTATGGTGGT 360

Qy        121 ProMetPheLeuMetProGlyLeuValIleThrLeuSerIleThrGlyAlaLeuAsnThr 140
              ||||||||||||:::||||||||:::||||||||||||||||||||||||||||||||||
Db        361 CCTATGTTTCTTCTCCCAGGACTGATAATTACACTCTCCATAACTGGAGCACTGAATACA 420

Qy        141 ValLeuThrGluGluHisArgLysGluIleCysArgTyrLeuTyrAsnHisGlnAsnLys 160
              ||||||:::|||:::|||::::::|||:::   ||||||||||||||||||||||:::
Db        421 GTATTGTCGGAACAACATAAACAAGAAATGCGCCGTTATCTCTATAATCACCAGAATGAG 480

Qy        161 AspGlyGlyTrpGlyLeuHisIleGluGlyProSerThrMetPheGlySerValLeuSer 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        481 GACGGAGGTTGGGGTTTACATATTGAGGGCCCTAGCACCATGTTTGGGTCTGTGTTGAAC 540

Qy        181 TyrIleThrLeuArgLeuLeuGlyGluGlyProAsnAspGlyGlnGlyGluMetGluLys 200
              |||:::||||||||||||||||||||||||||||||||||   |||:::||||||||||
Db        541 TATGTTACTCTAAGGTTGCTTGGAGAAGGACCTAACGATGGAGATGGAGATATGGAGAAA 600

Qy        201 AlaArgAspTrpIleLeuGlyHisGlyGlyAlaThrTyrIleThrSerTrpGlyLysMet 220
              |||||||||||||||   ||||||||||||||   ||||||||||||||||||||||
Db        601 GGACGAGACTGGATACTAAATCATGGTGGTGCTACCAATATTACATCTTGGGGGAAAATG 660

Qy        221 TrpLeuSerValLeuGlyValTyrGluTrpSerGlyAsnAsnProLeuProProGluIle 240
              ||||||||||||||||   ::||||||||||||||||||||||||||||||||||||||
Db        661 TGGCTATCGGTACTTGGAGCTTTTGAATGGTCCGGAAATAACCCACTGCCACCTGAGATA 720

Qy        241 TrpLeuLeuProTyrMetLeuProPheHisProGlyArgMetTrpCysHisCysArgMet 260
              ||||||||||||||||   ||||||   ||||||||||||||||||||||||||||||||
Db        721 TGGCTTCTCCCATATTTCCTGCCAATACATCCAGGAAGGATGTGGTGCCATTGTCGAATG 780

Qy        261 ValTyrLeuProMetSerTyrLeuTyrGlyLysArgPheValGlyProIleSerProThr 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::   |||
Db        781 GTGTACTTGCCGATGTCGTATTTGTATGGAAAAAGGTTTGTGGGTCCCATAACGTCCACT 840

Qy        281 ValLeuSerLeuArgLysGluLeuTyrThrValProTyrHisAspIleAspTrpAspGln 300
              ||||||||||||||||||||||:::|||||||||||||||:::::::::|||::::::
Db        841 GTTTTATCACTGAGAAAGGAGCTTTCACAGTACCATATCATGAAGTCAACTGGAATGAA 900

Qy        301 AlaArgAsnLeuCysAlaLysGluAspLeuTyrTyrProHisProLeuValGlnAspIle 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GCACGCAACCTTTGCGCAAAGGAGGATTTATACTACCCACATCCACTTGTGCAAGATATT 960

Qy        321 LeuTrpAlaSerLeuHisLysPheLeuGluProIleLeuMetHisTrpProGlyLysArg 340
              ||||||||||||||||||||   :::||||||:::||||||   |||||||||
Db        961 CTTTGGGCATCACTTCATAAGATTGTTGAGCCTGTTCTGATGCGATGGCCTGGTGCAAAT 1020

Qy        341 LeuArgGluLysAlaIleIleSerAlaLeuGluHisIleHisTyrGluAspGluAsnThr 360
              ||||||||||||||||||   :::|||:::||||||||||||||||||||||||||||||
Db       1021 TTGAGAGAAAAGGCTATAAGAACCGCAATAGAACATATTCATTATGAAGATGAGAATACT 1080

Qy        361 ArgTyrIleCysIleGlyProValAsnLysValLeuAsnMetLeuCysCysTrpValGlu 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AGGTACATCTGCATAGGTCCCGTGAACAAGGTATTAAATATGCTTTGCTGTTGGGTAGAA 1140

Qy        381 AspProAsnSerGluAlaPheLysLeuHisLeuProArgIleTyrAspTyrLeuTrpIle 400
              ||||||||||||||||||||||||||||||||||||||||||:::|||:::||||||:::
Db       1141 GACCCAAACTCAGAGGCTTTCAAGTTGCACCTACCAAGAATCCATGACTTTCTCTGGTTA 1200

Qy        401 AlaGluAspGlyMetLysMetGlnGlyTyrAsnGlySerGlnLeuTrpAspThrAlaPhe 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db       1201 GCTGAAGATGGAATGAAGATGCAGGGTTATAACGGAAGCCAGCTATGGGATACAGGTTTT 1260

Qy        421 AlaValGlnAlaIleIleAlaSerAsnLeuIleGluGluPheGlyProThrIleArgLys 440
              |||:::|||||||:::|||:::||||||:::|||||||   :::   |||
Db       1261 GCTATTCAAGCGATTTTGGCAACTAACCTCGTCGAAGAATATGGGCCCGTTTTGGAAAAA 1320
```

IQALISFRKLYPGHRKKEVDECIEKAVKFIESIQAADGSWYGSWAVCFTYGTWFGVKG
LVAVGKTLKNSPHVAKACEFLLSKQQPSGGWGESYLSCQDKVYSNLDGNRSHVVNTAW
AMLALIGAGQAEVDRKPLHRAARYLINAQMENGDFPQQEIMGVFNRNCMITYAAYRNI
FPIWALGEYRCQVLLQQGE"

ORIGIN

Alignment Scores:

Pred. No.:	1.48e-291	Length:	2561
Score:	3486.00	Matches:	608
Percent Similarity:	90.85%	Conservative:	77
Best Local Similarity:	80.64%	Mismatches:	69
Query Match:	84.04%	Indels:	0
DB:	8	Gaps:	0

US-09-419-679-6 (1-757) x U02555 (1-2561)



```
Db       1182 CTGAGAGAAAAGGCTCTTCGCACTGTGTTGGAACATATACACTATGAAGATGAAAATACT 1241

Qy        361 ArgTyrIleCysIleGlyProValAsnLysValLeuAsnMetLeuCysCysTrpValGlu 380

Db       1242 CGTTATATTTGCATAGGACCTGTAAACAAGGTGTTAAATATGCTCTGTTGTTGGGTGGAA 1301

Qy        381 AspProAsnSerGluAlaPheLysLeuHisLeuProArgIleTyrAspTyrLeuTrpIle 400

Db       1302 GATCCAAATTCCGAGGCTTTTAAGTTGCATCTTCCAAGAATAAATGACTATTTGTGGATT 1361

Qy        401 AlaGluAspGlyMetLysMetGlnGlyTyrAsnGlySerGlnLeuTrpAspThrAlaPhe 420

Db       1362 GCTGAAGATGGAATGAAAATGCAGGGTTATAATGGAAGTCAATTATGGGATACTGCCTTT 1421

Qy        421 AlaValGlnAlaIleIleAlaSerAsnLeuIleGluGluPheGlyProThrIleArgLys 440

Db       1422 GCTGTTCAAGCTATTATTTCAACCAACCTTTTTGAAGAATATGGTCCAACTCTCGAAAAA 1481

Qy        441 AlaHisThrTyrIleLysAsnSerGlnValLeuGluAspCysProGlyAspLeuAsnLys 460

Db       1482 GCACATATGTACATAAAAAAATCGCAGGTTCGAGAAGATTGCCCAGGTGATCTTGATTTT 1541

Qy        461 TrpTyrArgHisIleSerLysGlyAlaTrpProPheSerThrGlyAspHisGlyTrpPro 480

Db       1542 TGGTATCGCCACATTTCAAAAGGCGCTTGGCCTTTCTCAACTGCCGATCACGGATGGCCC 1601

Qy        481 IleSerAspCysThrAlaGluGlyLeuLysAlaValLeuLeuLeuSerLysIleAlaPro 500

Db       1602 ATATCAGACTGCACAGCAGAAGGATTGAAAGCTGCTCTTTTATTATCTAAAATTCCGCCA 1661

Qy        501 GluIleValGlyGluProIleAspValLysArgLeuTyrAspSerValAsnValIleLeu 520

Db       1662 GATGTTGTTGGTGAACCATTAGTTGAAGAGCGATTGTATGATGCTGTAAATGTCATTCTC 1721

Qy        521 SerLeuGlnAsnGluAspGlyGlyPheAlaThrTyrGluLeuLysArgSerTyrAsnTrp 540

Db       1722 TCATTACAGAATGCGGATGGTGGCTTTGCAACATATGAGCTCACAAGATCCTATCCCTGG 1781

Qy        541 LeuGluIleIleAsnProAlaGluThrPheGlyAspIleValIleAspTyrProTyrVal 560

Db       1782 TTGGAGTTGATCAACCCTGCTGAAACCTTTGGGGACATTGTTATTGATTATAATTATGTG 1841

Qy        561 GluCysThrSerAlaAlaIleGlnAlaLeuAlaSerPheArgLysLeuTyrProGlyHis 580

Db       1842 GAATGTACGTCAGCAGCAATTCAAGCTCTGACATCGTTTAAGAAATCATATCCCAAGCAT 1901

Qy        581 ArgArgGluGluIleGlnHisCysIleAspLysAlaThrThrPheIleGluLysIleGln 600

Db       1902 CGGGAGGAGGAGGTAGACGTCTGTATCAAAAGAGCTGCTATGTTCACTGAAAAGATACAA 1961

Qy        601 AlaSerAspGlySerTrpTyrGlySerTrpGlyValCysPheThrTyrGlyAlaTrpPhe 620

Db       1962 GCATCAGATGGTTCATGGTATGGTTCTTGGGGTGTTTGCTTCACTTACGGTACATGGTTT 2021

Qy        621 GlyValLysGlyLeuIleAlaAlaGlyArgSerPheSerAsnCysSerSerIleArgLys 640

Db       2022 GGGGTAAAAGGGTTAGTGGCTGCTGGAAAGAACTTCAATGATTGCTTTGGTATCCGCAAA 2081

Qy        641 AlaCysGluPheLeuLeuSerLysGlnLeuProSerGlyGlyTrpGlyGluSerTyrLeu 660

Db       2082 GCTTGTGATTTTCTGCTGTCTAAGCAGCTTCCTTCAGGTGGCTGGGGAGAGAGTTATCTT 2141

Qy        661 SerCysGlnAsnLysValTyrSerAsnLeuGluGlyAsnArgSerHisValValAsnThr 680

Db       2142 TCATGTCAAAACAAGGTGTATTCACATGTAGAGGGCAACAGGTCTCATGTTGTAAATACT 2201

Qy        681 GlyTrpAlaMetLeuAlaLeuIleAspAlaGlyGlnAlaLysArgAspSerGlnProLeu 700

Db       2202 GGATGGGCTATGTTGGCGCTTATAGAAGCTGGACAGGCTGAGAGAGACCCAACGCCATTG 2261

Qy        701 HisArgAlaAlaAlaTyrLeuIleAsnSerGlnLeuGluAspGlyAspPheProGlnGln 720

Db       2262 CACCGTGCAGCTAGGGTTTTGATAAATTCTCAAATGGAGAATGGAGATTTTCCTCAAGAG 2321

Qy        721 GluIleMetGlyValPheAsnLysAsnCysMetIleThrTyrAlaAlaTyrArgAsnIle 740

Db       2322 GAAATCATGGGGGTCTTCAATAGGAACTGTATGATTACCTACGCCGCTTACAGGAACATT 2381

Qy        741 PheProIleTrpAlaLeuGlyGluTyrGlnSerGlnValLeuGlnSer 756

Db       2382 TTTCCTATTTGGGCATTGGGAGAGTACCGCTGTCGGGTATTGCAGGCT 2429
```

```
RESULT 7
U02555
LOCUS       U02555                2561 bp    mRNA    linear   PLN 05-MAR-1998
DEFINITION  Arabidopsis thaliana Landsberg erecta cycloartenol synthase (CAS1)
            mRNA, complete cds.
ACCESSION   U02555 S67399
VERSION     U02555.1  GI:452445
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 2561)
  AUTHORS   Minet,M., Dufour,M.E. and Lacroute,F.
  TITLE     Complementation of Saccharomyces cerevisiae auxotrophic mutants by
            Arabidopsis thaliana cDNAs
  JOURNAL   Plant J. 2 (3), 417-422 (1992)
  MEDLINE   93272056
   PUBMED   1303803
REFERENCE   2  (bases 1 to 2561)
  AUTHORS   Corey,E.J., Matsuda,S.P. and Bartel,B.
  TITLE     Isolation of an Arabidopsis thaliana gene encoding cycloartenol
            synthase by functional expression in a yeast mutant lacking
            lanosterol synthase by the use of a chromatographic screen
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 90 (24), 11628-11632 (1993)
  MEDLINE   94089686
   PUBMED   7505443
REFERENCE   3  (bases 1 to 2561)
  AUTHORS   Matsuda,S.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-OCT-1993) Department of Chemistry, Harvard
            University, Cambridge, MA 02138, USA
COMMENT     On Mar 5, 1998 this sequence version replaced gi:455864.
FEATURES             Location/Qualifiers
     source          1. .2561
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /cultivar="Landsberg erecta ecotype"
                     /db_xref="taxon:3702"
                     /clone_lib="cDNA from Minet et al"
                     /dev_stage="seedling"
     gene            1. .2561
                     /gene="CAS1"
     CDS             54. .2333
                     /gene="CAS1"
                     /EC_number="5.4.99.8"
                     /function="cycloartenol synthesis; 2,3-oxidosqualene
                     metabolism"
                     /note="(S)-2,3-epoxysqualene mutase"
                     /codon_start=1
                     /product="cycloartenol synthase"
                     /protein_id="AAC04931.1"
                     /db_xref="GI:452446"
                     /translation="MWKLKIAEGGSPWLRTTNNHVGRQFWEFDPNLGTPEDLAAVEEA
                     RKSFSDNRFVQKHSADLLMRLQFSRENLISPVLPQVKIEDTDDVTEEMVETTLKRGLD
                     FYSTIQAHDGHWPGDYGGPMFLLPGLIITLSITGALNTVLSEQHKQEMRRYLYNHQNE
                     DGGWGLHIEGPSTMFGSVLNYVTLRLLGEGPNDGDGDMEKGRDWILNHGGATNITSWG
                     KMWLSVLGAFEWSGNNPLPPEIWLLPYFLPIHPGRMWCHCRMVYLPMSYLYGKRFVGP
                     ITSTVLSLRKELFTVPYHEVNWNEARNLCAKEDLYYPHPLVQDILWASLHKIVEPVLM
                     RWPGANLREKAIRTAIEHIHYEDENTRYICIGPVNKVLNMLCCWVEDPNSEAFKLHLP
                     RIHDFLWLAEDGMKMQGYNGSQLWDTGFAIQAILATNLVEEYGPVLEKAHSFVKNSQV
                     LEDCPGDLNYWYRHISKGAWPFSTADHGWPISDCTAEGLKAALLLSKVPKEIVGEPID
                     AKRLYEAVNVIISLQNADGGLATYELTRSYPWLELINPAETFGDIVIDYPYVECTSAA
```

```
Qy        721 GluIleMetGlyValPheAsnLysAsnCysMetIleThrTyrAlaAlaTyrArgAsnIle 740
              ||||||||||||||||||:::||||||||||||||||||||||||||||||||||||||||
Db       2245 GAGATCATGGGGGTCTTCGACAAGAATTGCATGATCACTTATGCTGCATACAGAAACATT 2304

Qy        741 PheProIleTrpAlaLeuGlyGluTyrGlnSerGlnValLeuGln 755
              ||||||||||||||||||||||||||||:::   :::|||||||||
Db       2305 TTCCCTATTTGGGCTTTGGGAGAATATCGATGTCGAGTACTCCAG 2349

RESULT 6
AB055510
LOCUS       AB055510               2514 bp    mRNA    linear   PLN 05-JAN-2002
DEFINITION  Betula platyphylla CASBPX2 mRNA for cycloartenol synthase, complete
            cds.
ACCESSION   AB055510
VERSION     AB055510.1  GI:18147591
KEYWORDS    .
SOURCE      Betula platyphylla
  ORGANISM  Betula platyphylla
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fagales; Betulaceae; Betula.
REFERENCE   1
  AUTHORS   Zhang,H., Shibuya,M., Ebizuka,Y. and Yokota,S.
  TITLE     Oxidosqualene Cyclases from Cell Suspension Culture of Betula
            platyphylla var. japonica:Molecular Evolution of Oxidosqualene
            Cyclases in Higher Plants
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2514)
  AUTHORS   Shibuya,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-FEB-2001) Masaaki Shibuya, University of Tokyo,
            Faculty of Pharmaceutical Sciences; 7-3-1 Hongo, Bunkyo-ku, Tokyo
            113-0033, Japan (E-mail:shibuyam@mol.f.u-tokyo.ac.jp,
            URL:http://www.f.u-tokyo.ac.jp/~tennen/head.htm,
            Tel:81-3-5841-4741, Fax:81-3-5841-4744)
FEATURES             Location/Qualifiers
     source          1. .2514
                     /organism="Betula platyphylla"
                     /mol_type="mRNA"
                     /db_xref="taxon:78630"
     gene            1. .2514
                     /gene="CASBPX2"
     CDS             162. .2435
                     /gene="CASBPX2"
                     /EC_number="5.4.99.8"
                     /codon_start=1
                     /product="cycloartenol synthase"
                     /protein_id="BAB83086.1"
                     /db_xref="GI:18147592"
                     /translation="MWKLKIAEGGSPWLRTLNNHVGRQVWEFDPKLGSPEELAEIERA
                     RETSLKVRFEKKHSSDLLMRIQFAKENPRGAVLPQVKVNETEDVTEEMVTRMLRRAIS
                     FHSTLQAHDGHWAGDYGGPMFLMPGLVITLSITGALNTVLSEEHKKEMCRYLYNHQNK
                     DGGWGLHIEGPSTMFGTVLSYVTLRLLGEGANDGQGAIERGRKWILDHGSATAIISWG
                     KMWLSVLGAFEWSGNNPLPPEIWLLPYMLPVHPGRMWCHCRMVYLPMSYLYGKRFVGP
                     ITPTVMSLRKELYSVPYHEIDWNQARNLCAKEXLYYPHPLVQDILWASLHKLVEPVLM
                     RWPGKRLREKALRTVLEHIHYEDENTRYICIGPVNKVLNMLCCWVEDPNSEAFKLHLP
                     RINDYLWIAEDGMKMQGYNGSQLWDTAFAVQAIISTNLFEEYGPTLEKAHMYIKKSQV
                     REDCPGDLDFWYRHISKGAWPFSTADHGWPISDCTAEGLKAALLLSKIPPDVVGEPLV
                     EERLYDAVNVILSLQNADGGFATYELTRSYPWLELINPAETFGDIVIDYNYVECTSAA
                     IQALTSFKKSYPKHREEEVDVCIKRAAMFTEKIQASDGSWYGSWGVCFTYGTWFGVKG
                     LVAAGKNFNDCFGIRKACDFLLSKQLPSGGWGESYLSCQNKVYSHVEGNRSHVVNTGW
                     AMLALIEAGQAERDPTPLHRAARVLINSQMENGDFPQEEIMGVFNRNCMITYAAYRNI
                     FPIWALGEYRCRVLQAP"
ORIGIN

Alignment Scores:
Pred. No.:                2.46e-294      Length:          2514
Score:                    3518.00        Matches:         627
Percent Similarity:       91.14%         Conservative:    62
Best Local Similarity:    82.94%         Mismatches:      67
Query Match:              84.81%         Indels:          0
DB:                       8              Gaps:            0
```

```
US-09-419-679-6 (1-757) x AB055510 (1-2514)

Qy          1 MetTrpLysLeuLysPheAlaGluGlyGlyAsnProTrpLeuArgThrLeuAsnAsnHis 20
              ||||||||||||||||   ||||||||||||:::||||||||||||||||||||||||||
Db        162 ATGTGGAAGCTGAAGATCGCGGAGGGAGGGAGTCCATGGCTGCGGACGTTGAACAATCAC 221

Qy         21 ValGlyArgGlnValTrpGluPheAspProLysLeuGlySerProGlnAspLeuLeuGlu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::::|||   |||
Db        222 GTCGGACGACAAGTTTGGGAATTCGATCCCAAGCTCGGGTCGCCTGAAGAGCTTGCCGAG 281

Qy         41 IleGluLysAlaArgGlnAsnPheHisAspAsnArgPheThrHisLysHisSerAlaAsp 60
              ||||||:::||||||:::                ||||||      |||||||||:::|||
Db        282 ATCGAAAGAGCTCGCGAGACTTCTCTGAAAGTCCGTTTCGAGAAAAAGCATAGCTCAGAT 341

Qy         61 LeuLeuMetArgMetGlnPheAlaArgGluAsnProThrArgGluValLeuProLysVal 80
              ||||||||||:::||||||||:::|||||||||          ||||||||||:::|||
Db        342 CTACTTATGCGGATTCAGTTCGCCAAGGAGAACCCACGTGGTGCAGTTCTACCCCAAGTC 401

Qy         81 GlyValLysAspIleGluAspValThrGlnGluIleValThrLysThrLeuArgArgAla 100
              |||   :::   ||||||||||||:::|||:::||||||:::   ||||||||||||
Db        402 AAAGTAAATGAAACAGAAGATGTTACAGAGGAGATGGTGACAAGGATGTTAAGAAGGGCT 461

Qy        101 ValSerPheHisSerThrLeuGlnCysHisAspGlyHisTrpProGlyAspTyrGlyGly 120
              :::||||||||||||||||||||||   ||||||||||||||   ||||||||||||||
Db        462 ATAAGTTTCCACTCAACTCTCCAGGCCCATGATGGACATTGGGCGGGGGACTATGGAGGT 521

Qy        121 ProMetPheLeuMetProGlyLeuValIleThrLeuSerIleThrGlyAlaLeuAsnThr 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        522 CCAATGTTTCTTATGCCTGGCTTGGTAATCACTCTGTCTATCACTGGAGCACTAAATACA 581

Qy        141 ValLeuThrGluGluHisArgLysGluIleCysArgTyrLeuTyrAsnHisGlnAsnLys 160
              ||||||:::||||||||:::|||||||:::||||||||||||||||||||||||||||||
Db        582 GTTTTATCTGAAGAACATAAAAAGGAAATGTGCCGTTACCTCTATAATCATCAGAACAAA 641

Qy        161 AspGlyGlyTrpGlyLeuHisIleGluGlyProSerThrMetPheGlySerValLeuSer 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db        642 GATGGTGGGTGGGGATTACACATTGAGGGCCCAAGTACCATGTTTGGTACTGTCTTAAGC 701

Qy        181 TyrIleThrLeuArgLeuLeuGlyGluGlyProAsnAspGlyGlnGlyGluMetGluLys 200
              |||:::||||||||||||||||||||   ||||||||||||||   :::|||:::
Db        702 TATGTTACGTTAAGGTTGCTTGGTGAAGGAGCTAATGATGGGCAAGGAGCAATTGAGAGA 761

Qy        201 AlaArgAspTrpIleLeuGlyHisGlyGlyAlaThrTyrIleThrSerTrpGlyLysMet 220
              |||   |||||||||   ||||||   ||||||   |||   ||||||||||||||||
Db        762 GGGCGTAAATGGATTCTGGACCATGGCAGTGCTACTGCAATTATATCATGGGGAAAGATG 821

Qy        221 TrpLeuSerValLeuGlyValTyrGluTrpSerGlyAsnAsnProLeuProProGluIle 240
              ||||||||||||||||||   :::||||||||||||||||||||||||||||||||||||
Db        822 TGGCTTTCAGTACTTGGAGCATTTGAATGGTCGGGCAATAATCCCCTTCCCCCTGAAATA 881

Qy        241 TrpLeuLeuProTyrMetLeuProPheHisProGlyArgMetTrpCysHisCysArgMet 260
              ||||||||||||||||||||||||   ||||||||||||||||||||||||||||||||||
Db        882 TGGCTTCTTCCTTACATGCTCCCTGTACATCCAGGAAGAATGTGGTGTCACTGTCGGATG 941

Qy        261 ValTyrLeuProMetSerTyrLeuTyrGlyLysArgPheValGlyProIleSerProThr 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        942 GTCTATTTGCCAATGTCTTACTTATATGGGAAGAGATTTGTTGGTCCAATCACGCCAACA 1001

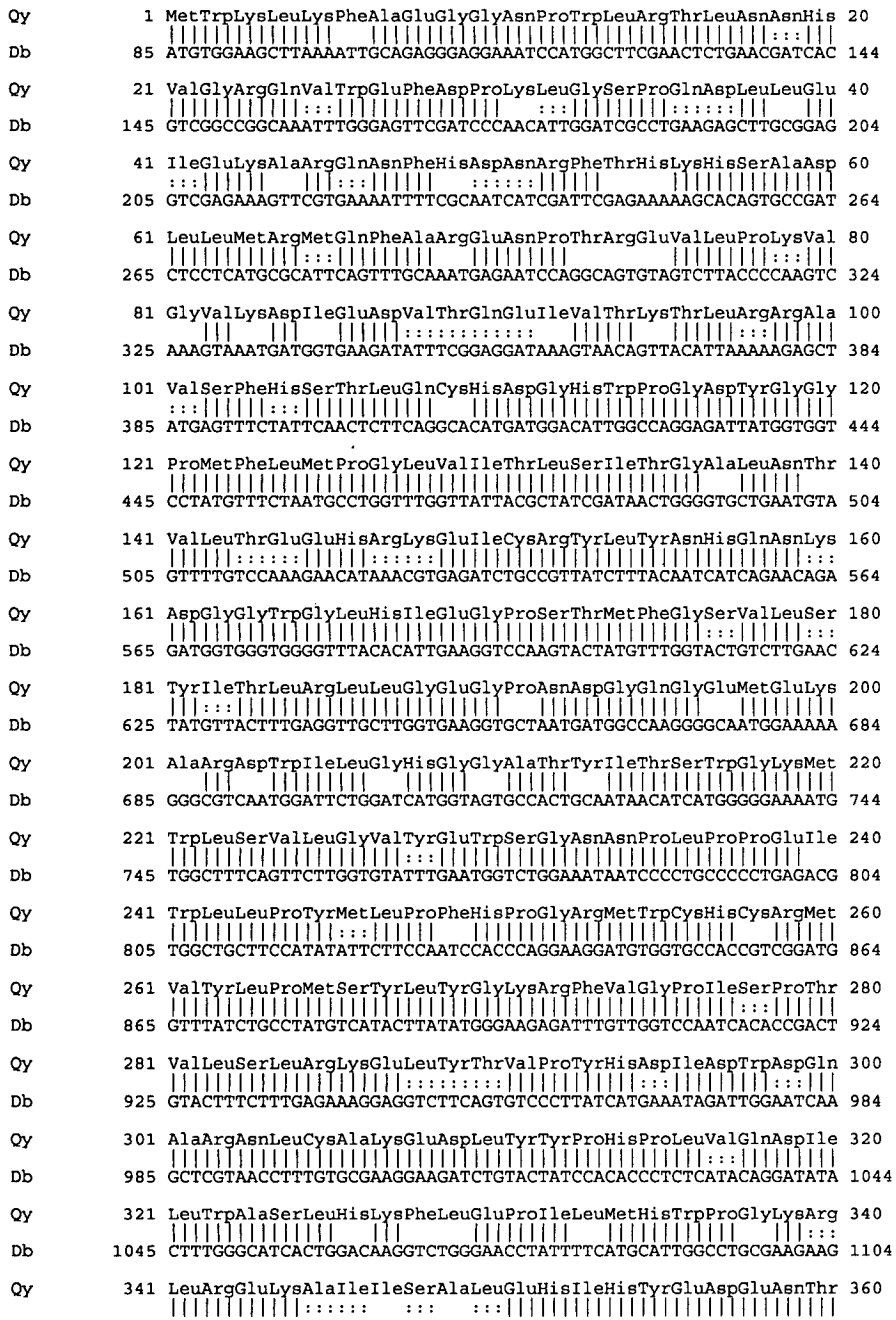
Qy        281 ValLeuSerLeuArgLysGluLeuTyrThrValProTyrHisAspIleAspTrpAspGln 300
              |||:::||||||||||||||||||||:::||||||||||||:::|||||||||:::|||
Db       1002 GTTATGTCTTTGAGGAAGGAGCTTTATTCTGTCCCCTATCATGAAATAGATTGGAATCAG 1061

Qy        301 AlaArgAsnLeuCysAlaLysGluAspLeuTyrTyrProHisProLeuValGlnAspIle 320
              ||||||||||||||||||||||||   ||||||||||||||||||||||||||||||||||
Db       1062 GCACGCAACCTATGTGCAAAGGAAKATCTGTACTACCCACATCCGCTGGTGCAGGATATA 1121

Qy        321 LeuTrpAlaSerLeuHisLysPheLeuGluProIleLeuMetHisTrpProGlyLysArg 340
              ||||||||||||||||||||||   :::||||||:::||||||   ||||||||||||||
Db       1122 CTTTGGGCATCTCTTCATAAGCTTGTTGAACCGGTTTTAATGCGTTGGCCTGGAAAAAGA 1181

Qy        341 LeuArgGluLysAlaIleIleSerAlaLeuGluHisIleHisTyrGluAspGluAsnThr 360
              ||||||||||||||:::   :::   ||||||||||||||||||||||||||||||||||
```

US-09-419-679-6 (1-757) x AB009029 (1-2559)



```
Qy        361 ArgTyrIleCysIleGlyProValAsnLysValLeuAsnMetLeuCysCysTrpValGlu 380
Db       1112 CGCTATATATGCATAGGGCCTGTAAACAAGGTGTTAAATATGCTATGCTGTTGGGCTGAG 1171

Qy        381 AspProAsnSerGluAlaPheLysLeuHisLeuProArgIleTyrAspTyrLeuTrpIle 400
Db       1172 GATCCAAACTCTGAGGCTTTCAAGCTGCATCTTCCAAGACTCTATGATTTTTTATGGCTT 1231

Qy        401 AlaGluAspGlyMetLysMetGlnGlyTyrAsnGlySerGlnLeuTrpAspThrAlaPhe 420
Db       1232 GCTGAGGATGGCATGAAAATGCAGGGGTACAATGGAAGTCAATTATGGGACACTGCTTTT 1291

Qy        421 AlaValGlnAlaIleIleAlaSerAsnLeuIleGluGluPheGlyProThrIleArgLys 440
Db       1292 GCTGTTCAAGCAATCATATCAACTAGTTTGACTGAATATTGTGGCCCAACTCTCAGAAAA 1351

Qy        441 AlaHisThrTyrIleLysAsnSerGlnValLeuGluAspCysProGlyAspLeuAsnLys 460
Db       1352 GCACATACATTCATGAAAGACTCACAGGTGCTAGATGATTGTCCTGGTGATCTTGATTTT 1411

Qy        461 TrpTyrArgHisIleSerLysGlyAlaTrpProPheSerThrGlyAspHisGlyTrpPro 480
Db       1412 TGGTATCGTCATATATCAAAAGGTGCATGGCCTTTCTCCACTGCTGATCATGGATGGCCC 1471

Qy        481 IleSerAspCysThrAlaGluGlyLeuLysAlaValLeuLeuLeuSerLysIleAlaPro 500
Db       1472 ATTTCAGATTGTACAGCCGAAGGATTTAAGGCTGTGCTTCTGTTATCAAAATTCCCAGCA 1531

Qy        501 GluIleValGlyGluProIleAspValLysArgLeuTyrAspSerValAsnValIleLeu 520
Db       1532 GAACTTGTTGGTGAACCATTAGATGCCAAGCGACTGTATGATGCTGTCAACGTTATTCTT 1591

Qy        521 SerLeuGlnAsnGluAspGlyGlyPheAlaThrTyrGluLeuLysArgSerTyrAsnTrp 540
Db       1592 TCTTTACAGAATAGTGATGGTGGCTATGCAACATATGAACTCACAAGATCCTATCGCTGG 1651

Qy        541 LeuGluIleIleAsnProAlaGluThrPheGlyAspIleValIleAspTyrProTyrVal 560
Db       1652 CTAGAGTTAATCAACCCTGCCGAGACGTTCGGGGACATTGTTATTGACTATCCTTATGCA 1711

Qy        561 GluCysThrSerAlaAlaIleGlnAlaLeuAlaSerPheArgLysLeuTyrProGlyHis 580
Db       1712 GAATGTACATCAGCTGCAATTCAAGCGTTATCAGCTTTCAAAAAATTATATCCTGGGCAC 1771

Qy        581 ArgArgGluGluIleGlnHisCysIleAspLysAlaThrThrPheIleGluLysIleGln 600
Db       1772 CGTCGGGAAGAGATACAGTGTTCTATTGAAAAAGCTGCAGACTTCATAGAGAAGATACAA 1831

Qy        601 AlaSerAspGlySerTrpTyrGlySerTrpGlyValCysPheThrTyrGlyAlaTrpPhe 620
Db       1832 GCTTCAGATGGCTCATGGTATGGCTCTTGGGGAGTTTGCTTCACCTATGGTACATGGTTT 1891

Qy        621 GlyValLysGlyLeuIleAlaAlaGlyArgSerPheSerAsnCysSerSerIleArgLys 640
Db       1892 GGTGTAAAAGGGTTGATATCTGCTGGGAGAACCTTCTCTAATTGCTCAAGCATTCGCAAG 1951

Qy        641 AlaCysGluPheLeuLeuSerLysGlnLeuProSerGlyGlyTrpGlyGluSerTyrLeu 660
Db       1952 GCTTGTTATTTTCTACTGTCCAAACAACTTGCTTCTGGTGGTTGGGGAGAGAGTTACCTC 2011

Qy        661 SerCysGlnAsnLysValTyrSerAsnLeuGluGlyAsnArgSerHisValValAsnThr 680
Db       2012 TCCTGTCAGAACAAGGTGTATACAAATCTTGAGGGAGAACGTTCTCATGTTGTAAATACT 2071

Qy        681 GlyTrpAlaMetLeuAlaLeuIleAspAlaGlyGlnAlaLysArgAspSerGlnProLeu 700
Db       2072 GGATGGGCTATGCTGGCTCTCATTGAGGCTGGGCAGGCAGAGAGAGATGCAACACCATTG 2131

Qy        701 HisArgAlaAlaAlaTyrLeuIleAsnSerGlnLeuGluAspGlyAspPheProGlnGln 720
Db       2132 CACCATGCAGCTAAATTATTGATTAATTCTCAAATGGAAAATGGAGATTTCCCTCAGGAG 2191

Qy        721 GluIleMetGlyValPheAsnLysAsnCysMetIleThrTyrAlaAlaTyrArgAsnIle 740
Db       2192 GAGATCATGGGGGTCTTCAACAAGAATTGTATGATCACTTATGCTGCGTACAGAAACATT 2251

Qy        741 PheProIleTrpAlaLeuGlyGluTyrGlnSerGlnValLeu 754
Db       2252 TTCCCTATTTGGGCTTTGGGGGACTACCGGTGTCGTGTACTT 2293
```

RESULT 5
AB009029
LOCUS AB009029 2559 bp mRNA linear PLN 03-OCT-1998
DEFINITION Panax ginseng OSCPNX1 mRNA for Cycloartenol Synthase, complete cds.
ACCESSION AB009029
VERSION AB009029.1 GI:3688597
KEYWORDS OSCPNX1; Cycloartenol Synthase.
SOURCE Panax ginseng
ORGANISM Panax ginseng
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; campanulids; Apiales; Araliaceae; Panax.
REFERENCE 1 (sites)
AUTHORS Kushiro,T., Shibuya,M. and Ebizuka,Y.
TITLE Beta-amyrin synthase--cloning of oxidosqualene cyclase that catalyzes the formation of the most popular triterpene among higher plants
JOURNAL Eur. J. Biochem. 256 (1), 238-244 (1998)
MEDLINE 98417451
PUBMED 9746369
REFERENCE 2 (bases 1 to 2559)
AUTHORS Kushiro,T.
TITLE Direct Submission
JOURNAL Submitted (20-NOV-1997) Tetsuo Kushiro, The University of Tokyo, Graduate School of Pharmaceutical Sciences; 7-3-1 Hongo, Bunkyo-ku, Tokyo 113, Japan (E-mail:tkushiro@mol.f.u-tokyo.ac.jp, Tel:03-3812-2111, Fax:03-5689-2962)
COMMENT Sequence updated (15-May-1998).
FEATURES Location/Qualifiers
source 1. .2559
/organism="Panax ginseng"
/mol_type="mRNA"
/db_xref="taxon:4054"
/tissue_type="Hairy Root"
gene 1. .2559
/gene="OSCPNX1"
CDS 85. .2361
/gene="OSCPNX1"
/codon_start=1
/product="Cycloartenol Synthase"
/protein_id="BAA33460.1"
/db_xref="GI:3688598"
/translation="MWKLKIAEGDNPWLRTLNDHVGRQIWEFDPNIGSPEELAEVEKV
RENFRNHRFEKKHSADLLMRIQFANENPGSVVLPQVKVNDGEDISEDKVTVTLKRAMS
FYSTLQAHDGHWPGDYGGPMFLMPGLVITLSITGVLNVVLSKEHKREICRYLYNHQNR
DGGWGLHIEGPSTMFGTVLNYVTLRLLGEGANDGQGAMEKGRQWILDHGSATAITSWG
KMWLSVLGVFEWSGNNPLPPETWLLPYILPIHPGRMWCHRRMVYLPMSYLYGKRFVGP
ITPTVLSLRKEVFSVPYHEIDWNQARNLCAKEDLYYPHPLIQDILWASLDKVWEPIFM
HWPAKKLREKSLRTVMEHIHYEDENTRYICIGPVNKVLNMLCCWVEDPNSEAFKLHLP
RLHDFLWLAEDGMKMQGYNGSQLWDTAFAVQAIISTNLAEEYGPTLRKAHTFMKNSQV
LDDCPGDLDAWYRHVSKGAWPFSTADHGWPISDCTAEGFKAVLQLSKLPSELVGEPLD
AKRLYDAVNVILSLQNSDGGYATYELTRSYSWLELVNPAETFGDIVIDYPYVECTSAA
IQALTAFKKLFPGHRREEIQHSIEKAALFIEKIQSSDGSWYGSWGVCFTYGTWFGIKG
LVTAGRTFSSCASIRKACDFLLSKQVASGGWGESYLSCQNKVYTNLEGNRSHVVNTGW
AMLALIDAGQAERDATPLHRAAKLLINSQMENGDFPQEEIMGVFDKNCMITYAAYRNI
FPIWALGEYRCRVLQGPS"
ORIGIN

Alignment Scores:

Pred. No.:	3.42e-295	Length:	2559
Score:	3528.00	Matches:	616
Percent Similarity:	92.45%	Conservative:	82
Best Local Similarity:	81.59%	Mismatches:	57
Query Match:	85.05%	Indels:	0
DB:	8	Gaps:	0

```
Qy        701 HisArgAlaAlaAlaTyrLeuIleAsnSerGlnLeuGluAspGlyAspPheProGlnGln 720
              ||||||||||||     ||||||||||||||||||||:::||||||||||||||:::
Db       2216 CACCGTGCAGCTGTATGCTTGATAAATTCCCAATTGGAGAATGGTGACTTTCCGCAAGAG 2275

Qy        721 GluIleMetGlyValPheAsnLysAsnCysMetIleThrTyrAlaAlaTyrArgAsnIle 740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db       2276 GAAATAATGGGAGTCTTCAACAAGAATTGCATGATCACTTATGCTGCATACAGATGCATA 2335

Qy        741 PheProIleTrpAlaLeuGlyGluTyrGlnSerGlnValLeuGlnSer 756
              ||||||||||||||||||||||||||:::   :::||||||||||:::
Db       2336 TTCCCCATATGGGCGTTAGGAGAATACCGC---CGTGTATTGCAGGCC 2380

RESULT 4
AY520819
LOCUS       AY520819                2547 bp    mRNA    linear   PLN 03-FEB-2004
DEFINITION  Centella asiatica cycloartenol synthase (OSCCCS) mRNA, complete
            cds.
ACCESSION   AY520819
VERSION     AY520819.1  GI:41387167
KEYWORDS    .
SOURCE      Centella asiatica
  ORGANISM  Centella asiatica
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; campanulids; Apiales; Mackinlayaceae; Centella.
REFERENCE   1  (bases 1 to 2547)
  AUTHORS   Kim,O.T., Kim,M.Y., Ahn,J.C. and Hwang,B.
  TITLE     Molecular cloning and characterization of cDNAs encoding
            oxidosqualene cyclase from Centella asiatica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2547)
  AUTHORS   Kim,O.T., Kim,M.Y., Ahn,J.C. and Hwang,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JAN-2004) Department of Biology, Chonnam National
            University, Gwangju 500-757, Korea
FEATURES             Location/Qualifiers
     source          1. .2547
                     /organism="Centella asiatica"
                     /mol_type="mRNA"
                     /db_xref="taxon:48106"
     gene            1. .2547
                     /gene="OSCCCS"
     CDS             32. .2305
                     /gene="OSCCCS"
                     /note="oxidosqualene cyclase"
                     /codon_start=1
                     /product="cycloartenol synthase"
                     /protein_id="AAS01524.1"
                     /db_xref="GI:41387168"
                     /translation="MWKLKVAEGGNPWLRTVNNHVGRQIWEFDPKLGSPEELAEIEKA
                     RENFHNHRFEKQHSSDLLMRLQFANENPRHEVLPQVKVKDIGDISEDKVTITLKRALS
                     FYSTLQANDGHWAGDYGGPMFLMPGLVITLSITEALNAILSKEHKREICRYLYNHQNR
                     DGGWGLHIEGPSTMFGSVLNYVTLRLLGEGTKDGQGAMEKGRQWILDHGGATAITSWG
                     RMWLSVLGVFEWSGNNPLPPEIWLFPYNLPFHPGRMWCHCRMVYLPMSYLYGKRFVGP
                     ITPTVLSLRKELFTVPYHEIDWNQARNLCAKEDLYYPHPLVQDILWASIDKVLEPILM
                     RWPGKKLREKALRTVMEHIHYEDENTRYICIGPVNKVLNMLCCWAEDPNSEAFKLHLP
                     RLYDFLWLAEDGMKMQGYNGSQLWDTAFAVQAIISTSLTEYCGPTLRKAHTFMKDSQV
                     LDDCPGDLDFWYRHISKGAWPFSTADHGWPISDCTAEGFKAVLLLSKFPAELVGEPLD
                     AKRLYDAVNVILSLQNSDGGYATYELTRSYRWLELINPAETFGDIVIDYPYAECTSAA
                     IQALSAFKKLYPGHRREEIQCSIEKAADFIEKIQASDGSWYGSWGVCFTYGTWFGVKG
                     LISAGRTFSNCSSIRKACYFLLSKQLASGGWGESYLSCQNKVYTNLEGERSHVVNTGW
                     AMLALIEAGQAERDATPLHHAAKLLINSQMENGDFPQEEIMGVFNKNCMITYAAYRNI
                     FPIWALGDYRCRVLLAP"
ORIGIN

Alignment Scores:
Pred. No.:              5.66e-296       Length:         2547
Score:                  3537.00         Matches:        626
Percent Similarity:     92.44%          Conservative:   71
Best Local Similarity:  83.02%          Mismatches:     57
Query Match:            85.27%          Indels:         0
DB:                     8               Gaps:           0
```

```
US-09-419-679-6 (1-757) x AY520819 (1-2547)

Qy          1 MetTrpLysLeuLysPheAlaGluGlyGlyAsnProTrpLeuArgThrLeuAsnAsnHis 20
              ||||||||||||||   ||||||||||||||||||||||||||||||:::||||||||
Db         32 ATGTGGAAGCTCAAAGTCGCTGAAGGAGGAAATCCATGGCTTCGGACTGTCAACAATCAC 91

Qy         21 ValGlyArgGlnValTrpGluPheAspProLysLeuGlySerProGlnAspLeuLeuGlu 40
              ||||||||||||:::|||||||||||||||||||||||||||||:::|||   |||
Db         92 GTCGGCCGGCAAATTTGGGAGTTCGATCCAAAACTTGGATCTCCCGAAGAGCTTGCGGAG 151

Qy         41 IleGluLysAlaArgGlnAsnPheHisAspAsnArgPheThrHisLysHisSerAlaAsp 60
              ||||||||||||||:::|||||||||:::::|||||      :::|||||:::|||
Db        152 ATCGAGAAAGCTCGCGAGAATTTTCACAATCATCGATTCGAGAAACAGCACAGTTCCGAT 211

Qy         61 LeuLeuMetArgMetGlnPheAlaArgGluAsnProThrArgGluValLeuProLysVal 80
              ||||||||||:::|||||||||   |||||||||      ||||||||||||:::|||
Db        212 CTACTGATGCGGCTACAGTTCGCCAATGAGAATCCACGCCATGAAGTCTTACCCCAAGTT 271

Qy         81 GlyValLysAspIleGluAspValThrGlnGluIleValThrLysThrLeuArgArgAla 100
              ||||||||||||   |||:::::::::::   ||||||   ||||||:::||||||
Db        272 AAAGTAAAAGATATCGGAGATATTTCGGAGGATAAAGTAACAATTACGTTAAAAAGAGCT 331

Qy        101 ValSerPheHisSerThrLeuGlnCysHisAspGlyHisTrpProGlyAspTyrGlyGly 120
              :::||||||:::|||||||||||||   :::|||||||||||||   ||||||||||||||
Db        332 TTGAGTTTTTATTCAACTCTTCAAGCCAATGATGGACATTGGGCAGGAGATTATGGTGGT 391

Qy        121 ProMetPheLeuMetProGlyLeuValIleThrLeuSerIleThrGlyAlaLeuAsnThr 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db        392 CCTATGTTTTTAATGCCTGGTTTGGTTATCACTCTATCAATTACTGAGGCACTGAATGCT 451

Qy        141 ValLeuThrGluGluHisArgLysGluIleCysArgTyrLeuTyrAsnHisGlnAsnLys 160
              :::|||::::::||||||::::::||||||||||||||||||||||||||||||||:::
Db        452 ATCTTGTCCAAAGAACACAAACGTGAGATATGCCGCTATCTTTACAATCATCAGAATAGA 511

Qy        161 AspGlyGlyTrpGlyLeuHisIleGluGlyProSerThrMetPheGlySerValLeuSer 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        512 GATGGTGGGTGGGGTTTACACATTGAAGGTCCAAGCACTATGTTTGGTAGTGTCTTGAAC 571

Qy        181 TyrIleThrLeuArgLeuLeuGlyGluGlyProAsnAspGlyGlnGlyGluMetGluLys 200
              |||:::|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        572 TATGTTACTCTGAGATTGCTTGGTGAAGGAACTAAGGATGGGCAAGGAGCAATGGAGAAA 631

Qy        201 AlaArgAspTrpIleLeuGlyHisGlyGlyAlaThrTyrIleThrSerTrpGlyLysMet 220
              |||   |||||||||   ||||||||||||||||   ||||||||||||||:::|||
Db        632 GGGCGTCAGTGGATTCTGGATCATGGTGGTGCCACTGCAATTACATCATGGGGAAGAATG 691

Qy        221 TrpLeuSerValLeuGlyValTyrGluTrpSerGlyAsnAsnProLeuProProGluIle 240
              ||||||||||||||||||:::||||||||||||||||||||||||||||||||||||||
Db        692 TGGCTTTCAGTTCTTGGTGTATTTGAATGGTCTGGAAATAATCCCCTCCCCCCTGAGATA 751

Qy        241 TrpLeuLeuProTyrMetLeuProPheHisProGlyArgMetTrpCysHisCysArgMet 260
              ||||||   ||||||   ||||||||||||||||||||||||||||||||||||||||||
Db        752 TGGTTATTTCCATATAATCTTCCATTCCACCCAGGAAGGATGTGGTGCCACTGTCGGATG 811

Qy        261 ValTyrLeuProMetSerTyrLeuTyrGlyLysArgPheValGlyProIleSerProThr 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        812 GTTTATCTTCCTATGTCATACTTATATGGGAAGAGGTTTGTTGGTCCAATCACACCGACT 871

Qy        281 ValLeuSerLeuArgLysGluLeuTyrThrValProTyrHisAspIleAspTrpAspGln 300
              ||||||||||||||||||||||:::|||||||||||||||:::|||||||||:::|||
Db        872 GTCCTTTCTTTGAGAAAGGAGCTCTTCACTGTCCCTTATCATGAAATAGATTGGAATCAA 931

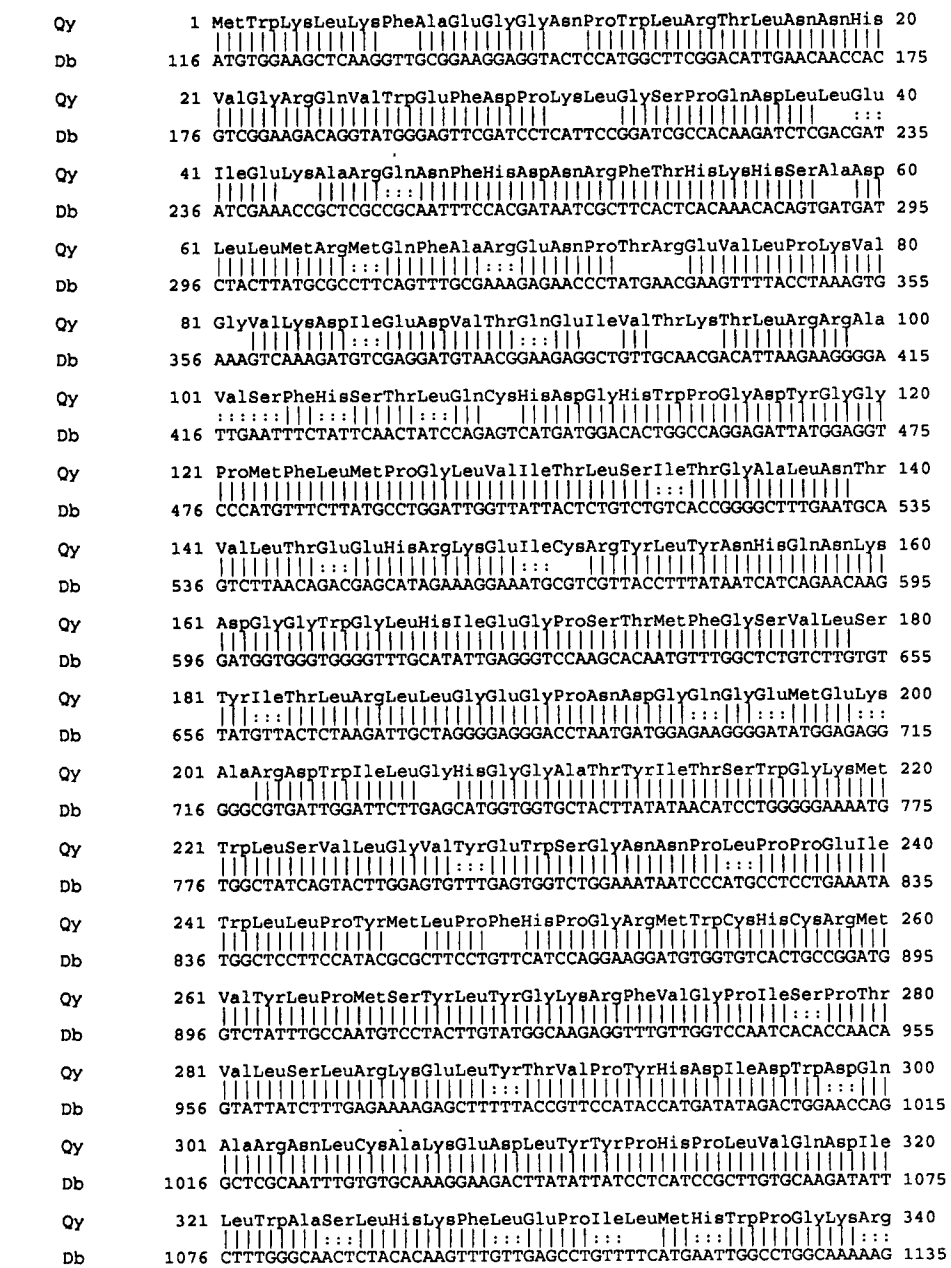
Qy        301 AlaArgAsnLeuCysAlaLysGluAspLeuTyrTyrProHisProLeuValGlnAspIle 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        932 GCACGCAACCTATGTGCAAAGGAAGATCTGTACTATCCACACCCTCTTGTGCAGGATATA 991

Qy        321 LeuTrpAlaSerLeuHisLysPheLeuGluProIleLeuMetHisTrpProGlyLysArg 340
              ||||||||||||:::   |||   ||||||||||||||||   ||||||||||||:::
Db        992 CTTTGGGCATCAATTGATAAGGTTTTGGAACCTATTTTAATGCGTTGGCCTGGAAAAAAG 1051

Qy        341 LeuArgGluLysAlaIleIleSerAlaLeuGluHisIleHisTyrGluAspGluAsnThr 360
              ||||||||||||||:::   :::   :::||||||||||||||||||||||||||||||
Db       1052 TTGAGAGAGAAAGCTCTTCGCACCGTGATGGAACACATACATTATGAAGATGAAAATACT 1111
```

Best Local Similarity: 87.04% Mismatches: 46
Query Match: 88.68% Indels: 1
DB: 8 Gaps: 1

US-09-419-679-6 (1-757) x D89619 (1-2677)



```
Qy        341 LeuArgGluLysAlaIleIleSerAlaLeuGluHisIleHisTyrGluAspGluAsnThr 360
Db       1151 TTGAGAGAAATGGCTATTAAGACTGCAATCGAGCATATACATTATGAAGATGATAATACT 1210

Qy        361 ArgTyrIleCysIleGlyProValAsnLysValLeuAsnMetLeuCysCysTrpValGlu 380
Db       1211 CGTTATCTTTGCATAGGTCCTGTAAATAAGGTGTTAAATATGCTTTGCTGTTGGGTGGAG 1270

Qy        381 AspProAsnSerGluAlaPheLysLeuHisLeuProArgIleTyrAspTyrLeuTrpIle 400
Db       1271 GATCCAAATTCAGAGGCCTTCAAGTTGCATCTTCCAAGGATCTATGATTATCTATGGATT 1330

Qy        401 AlaGluAspGlyMetLysMetGlnGlyTyrAsnGlySerGlnLeuTrpAspThrAlaPhe 420
Db       1331 GCTGAAGATGGCATGAAAATGCAGGGCTACAATGGAAGTCAACTATGGGATACTGCATTT 1390

Qy        421 AlaValGlnAlaIleIleAlaSerAsnLeuIleGluGluPheGlyProThrIleArgLys 440
Db       1391 ACTGCTCAAGCAATTATTTCATCTAACCTCATCGAAGAATATGGTCCAACTCTAAGAAAA 1450

Qy        441 AlaHisThrTyrIleLysAsnSerGlnValLeuGluAspCysProGlyAspLeuAsnLys 460
Db       1451 GCCCATACATACATTAAGAATTCGCAGGTTTTAGAAGATTGTCCAGGTGATCTTAGTAAA 1510

Qy        461 TrpTyrArgHisIleSerLysGlyAlaTrpProPheSerThrGlyAspHisGlyTrpPro 480
Db       1511 TGGTATCGTCACATTTCAAAAGGTGCTTGGCCTTTTTCGACTGCAGATCATGGATGGCCA 1570

Qy        481 IleSerAspCysThrAlaGluGlyLeuLysAlaValLeuLeuLeuSerLysIleAlaPro 500
Db       1571 ATTTCTGACTGCACAGCTGAAGGACTGAAAGCTGTTCTATTGCTATCTAAAATTGCACCA 1630

Qy        501 GluIleValGlyGluProIleAspValLysArgLeuTyrAspSerValAsnValIleLeu 520
Db       1631 GAGATTGTTGGTGAGCCATTGGATGCAAAGCGATTATATGATGCTGTAAATGTCATTCTC 1690

Qy        521 SerLeuGlnAsnGluAspGlyGlyPheAlaThrTyrGluLeuLysArgSerTyrAsnTrp 540
Db       1691 TCATTACAGAATGAAGATGGTGGCTTTGCAACATATGAGCTCACAAGATCTTATACCTGG 1750

Qy        541 LeuGluIleIleAsnProAlaGluThrPheGlyAspIleValIleAspTyrProTyrVal 560
Db       1751 TTGGAGTTAATCAATCCGGCTGAGACTTTTGGTGACATCGTTATTGATTACCCTTATGTG 1810

Qy        561 GluCysThrSerAlaAlaIleGlnAlaLeuAlaSerPheArgLysLeuTyrProGlyHis 580
Db       1811 GAATGTACCTCAGCGGCAATTCAAGCTTTGACATCATTTAAGAAACTATATCCTGGGCAT 1870

Qy        581 ArgArgGluGluIleGlnHisCysIleAspLysAlaThrThrPheIleGluLysIleGln 600
Db       1871 CGTCGAGAAGAAATACAATGCTGTATTGAAAAAGCTGCTTCCTTTATTGAAAAAACACAA 1930

Qy        601 AlaSerAspGlySerTrpTyrGlySerTrpGlyValCysPheThrTyrGlyAlaTrpPhe 620
Db       1931 GCTTCAGATGGTTCATGGTATGGTTCTTGGGGAGTTTGCTTCACTTATGGCACTTGGTTT 1990

Qy        621 GlyValLysGlyLeuIleAlaAlaGlyArgSerPheSerAsnCysSerSerIleArgLys 640
Db       1991 GGGGTAAAAGGGCTGATTGCTGCTGGAAAGAGTTTCAATAATTGCTCAAGCATCCGTAAA 2050

Qy        641 AlaCysGluPheLeuLeuSerLysGlnLeuProSerGlyGlyTrpGlyGluSerTyrLeu 660
Db       2051 GCTTGTGAATTTCTGCTGTCTAAGCAGCTCCCATCTGGTGGCTGGGGAGAGAGCTATCTG 2110

Qy        661 SerCysGlnAsnLysValTyrSerAsnLeuGluGlyAsnArgSerHisValValAsnThr 680
Db       2111 TCATGCCAAAACAAGGTGTATTCAAATGTGGAAAGCAACCGGTCTCATGTAGTAAATACT 2170

Qy        681 GlyTrpAlaMetLeuAlaLeuIleAspAlaGlyGlnAlaLysArgAspSerGlnProLeu 700
Db       2171 GGGTGGGCTATGCTGGCTCTCATTGATGCTGAACAGGCTAAGAGAGATCCAACGCCACTG 2230

Qy        701 HisArgAlaAlaAlaTyrLeuIleAsnSerGlnLeuGluAspGlyAspPheProGlnGln 720
Db       2231 CACCGTGCAGCTGTATACTTGATAAATTCCCAAATGGAGAATGGTGACTTTCCGCAACAG 2290

Qy        721 GluIleMetGlyValPheAsnLysAsnCysMetIleThrTyrAlaAlaTyrArgAsnIle 740
Db       2291 GAAATAATGGGAGTCTTCAACAAGAATTGCATGATCACATACGCAGCATACAGAAATGTA 2350

Qy        741 PheProIleTrpAlaLeuGlyGluTyrGlnSerGlnValLeuGlnSerArg 757
Db       2351 TTCCCCATATGGGCGTTGGGAGAATACCGTCATCGCGTATTGCAGTCGCAG 2401
```

RESULT 3
D89619
LOCUS D89619 2677 bp mRNA linear PLN 14-JAN-2000
DEFINITION Pisum sativum mRNA for cycloartenol synthase, complete cds.
ACCESSION D89619
VERSION D89619.1 GI:2627180
KEYWORDS cycloartenol synthase.
SOURCE Pisum sativum (pea)
ORGANISM Pisum sativum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Vicieae; Pisum.
REFERENCE 1 (sites)
AUTHORS Morita,M., Shibuya,M., Lee,M.S., Sankawa,U. and Ebizuka,Y.
TITLE Molecular cloning of pea cDNA encoding cycloartenol synthase and its functional expression in yeast
JOURNAL Biol. Pharm. Bull. 20 (7), 770-775 (1997)
MEDLINE 97399291
PUBMED 9255418
REFERENCE 2 (bases 1 to 2677)
AUTHORS Shibuya,M.
TITLE Direct Submission
JOURNAL Submitted (28-NOV-1996) Masaaki Shibuya, The University of Tokyo, Faculty of Pharmaceutical Sciences; 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan (E-mail:shibuyam@mol.f.u-tokyo.ac.jp, Tel:81-3-5841-4741, Fax:81-3-5841-4744)

FEATURES Location/Qualifiers
source 1. .2677
/organism="Pisum sativum"
/mol_type="mRNA"
/db_xref="taxon:3888"
/dev_stage="germination"
gene 1. .2677
/gene="CASPEA"
CDS 116. .2386
/gene="CASPEA"
/EC_number="5.4.99.8"
/codon_start=1
/product="cycloartenol synthase"
/protein_id="BAA23533.1"
/db_xref="GI:2627181"
/translation="MWKLKVAEGGTPWLRTLNNHVGRQVWEFDPHSGSPQDLDDIETA
RRNFHDNRFTHKHSDDLLMRLQFAKENPMNEVLPKVKVKDVEDVTEEAVATTLRRGLN
FYSTIQSHDGHWPGDYGGPMFLMPGLVITLSVTGALNAVLTDEHRKEMRRYLYNHQNK
DGGWGLHIEGPSTMFGSVLCYVTLRLLGEGPNDGEGDMERGRDWILEHGGATYITSWG
KMWLSVLGVFEWSGNNPMPPEIWLLPYALPVHPGRMWCHCRMVYLPMSYLYGKRFVGP
ITPTVLSLRKELFTVPYHDIDWNQARNLCAKEDLYYPHPLVQDILWATLHKFVEPVFM
NWPGKKLREKAIKTAIEHIHYEDENTRYICIGPVNKVLNMLCCWVEDPNSEAFKLHLP
RIYDYLWVAEDGMKMQGYNGSQLWDTAFAAQAIISTNLIDEFGPTLKKAHAFIKNSQV
SEDCPGDLSKWYRHISKGAWPFSTADHGWPISDCTAEGLKAVLLLSKIAPEIVGEPLD
SKRLYDAVNVILSLQNENGGLATYELTRSYTWLEIINPAETFGDIVIDCPYVECTSAA
IQALATFGKLYPGHRREEIQCCIEKAVAFIEKIQASDGSWYGSWGVCFTYGTWFGIKG
LIAAGKNFSNCLSIRKACEFLLSKQLPSGGWAESYLSCQNKVYSNLEGNRSHVVNTGW
AMLALIEAEQAKRDPTPLHRAAVCLINSQLENGDFPQEEIMGVFNKNCMITYAAYRCI
FPIWALGEYRRVLQAC"

ORIGIN

Alignment Scores:
Pred. No.: 3.41e-308 Length: 2677
Score: 3678.50 Matches: 658
Percent Similarity: 93.78% Conservative: 51

```
Qy       721 GluIleMetGlyValPheAsnLysAsnCysMetIleThrTyrAlaAlaTyrArgAsnIle 740
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2185 GAAATAATGGGAGTCTTCAACAAGAATTGCATGATCACATACGCCGCATACAGAAACATA 2244

Qy       741 PheProIleTrpAlaLeuGlyGluTyrGlnSerGlnValLeuGlnSerArg 757
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2245 TTCCCCATTTGGGCGTTGGGAGAATACCAATCCCAAGTATTGCAATCTCGT 2295
```

RESULT 2
AB025968
LOCUS AB025968 2889 bp mRNA linear PLN 14-DEC-2001
DEFINITION Glycyrrhiza glabra mRNA for cycloartenol synthase, complete cds.
ACCESSION AB025968
VERSION AB025968.1 GI:4589851
KEYWORDS cycloartenol synthase.
SOURCE Glycyrrhiza glabra
ORGANISM Glycyrrhiza glabra
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Galegeae; Glycyrrhiza.
REFERENCE 1
AUTHORS Hayashi,H., Hiraoka,N., Ikeshiro,Y., Kushiro,T., Morita,M., Shibuya,M. and Ebizuka,Y.
TITLE Molecular cloning and characterization of a cDNA for Glycyrrhiza glabra cycloartenol synthase
JOURNAL Biol. Pharm. Bull. 23 (2), 231-234 (2000)
MEDLINE 20168908
PUBMED 10706391
REFERENCE 2 (bases 1 to 2889)
AUTHORS Hayashi,H.
TITLE Direct Submission
JOURNAL Submitted (08-APR-1999) Hiroaki Hayashi, Gifu Pharmaceutical University, Department of Pharmacognosy; 5-6-1 Mitahora-higashi, Gifu, Gifu 502-8585, Japan (E-mail:hayashih@gifu-pu.ac.jp, Tel:81-58-237-3931(ex.325), Fax:81-58-237-5979)

FEATURES Location/Qualifiers
source 1. .2889
/organism="Glycyrrhiza glabra"
/mol_type="mRNA"
/db_xref="taxon:49827"
/clone="GgCAS1"
/cell_type="suspension cultured cells"
/clone_lib="lambda ZAPII"
gene 1. .2889
/gene="GgCAS1"
CDS 131. .2404
/gene="GgCAS1"
/codon_start=1
/product="cycloartenol synthase"
/protein_id="BAA76902.1"
/db_xref="GI:4589852"
/translation="MWKLKIAEGGSPWLRTVNNHVGRQVWEFDPKLGSPEDLLEIEKA
RQNFHDNRFTHKHSADLLMRIHFAKENPMNEVLPKVRVKDIEDVTEETVKTTLRRAIN
FHSTLQSHDGHWPGDYGGPMFLMPGLVITLSITGALNAVLTEEHRKEICRYLYNHQNK
DGGWGLHIEGPSTMFGSVLNYVALRLLGEGPNDRQGEMEKGRDWILGHGGATFITSWG
KMWLSVLGVYEWSGNNPLPPEIWLLPYVLPIHPGRMWCHCRMVYLPMSYLYGKRFVGP
ITPTILSLRKELYTIPYHDIDWNQARNLCAKEDLYYPHPLVQDILWASLHKFLEPILM
HWPGKKLREMAIKTAIEHIHYEDDNTRYLCIGPVNKVLNMLCCWVEDPNSEAFKLHLP
RIYDYLWIAEDGMKMQGYNGSQLWDTAFTAQAIISSNLIEEYGPTLRKAHTYIKNSQV
LEDCPGDLSKWYRHISKGAWPFSTADHGWPISDCTAEGLKAVLLLSKIAPEIVGEPLD
AKRLYDAVNVILSLQNEDGGFATYELTRSYTWLELINPAETFGDIVIDYPYVECTSAA
IQALTSFKKLYPGHRREEIQCCIEKAASFIEKTQASDGSWYGSWGVCFTYGTWFGVKG
LIAAGKSFNNCSSIRKACEFLLSKQLPSGGWGESYLSCQNKVYSNVESNRSHVVNTGW
AMLALIDAEQAKRDPTPLHRAAVYLINSQMENGDFPQQEIMGVFNKNCMITYAAYRNV
FPIWALGEYRHRVLQSQ"

ORIGIN

Alignment Scores:
Pred. No.: 0 Length: 2889
Score: 3837.00 Matches: 684
Percent Similarity: 95.64% Conservative: 40
Best Local Similarity: 90.36% Mismatches: 33

Query Match: 92.50% Indels: 0
DB: 8 Gaps: 0

US-09-419-679-6 (1-757) x AB025968 (1-2889)

```
Qy         1 MetTrpLysLeuLysPheAlaGluGlyGlyAsnProTrpLeuArgThrLeuAsnAsnHis 20
             |||||||||||||||  |||||||||||:||||||||||||||:|||||||||
Db       131 ATGTGGAAGCTCAAGATTGCGGAGGGAGGGAGTCCATGGCTTCGGACAGTGAACAATCAC 190

Qy        21 ValGlyArgGlnValTrpGluPheAspProLysLeuGlySerProGlnAspLeuLeuGlu 40
             ||||||||||||||||||||||||||||||||||||||||||:||||||||||||||
Db       191 GTTGGAAGACAGGTGTGGGAGTTCGATCCTAAGCTTGGATCGCCAGAAGATCTCCTCGAG 250

Qy        41 IleGluLysAlaArgGlnAsnPheHisAspAsnArgPheThrHisLysHisSerAlaAsp 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       251 ATTGAAAAGGCTCGCCAGAATTTCCACGACAACCGCTTCACTCACAAACACAGCGCTGAT 310

Qy        61 LeuLeuMetArgMetGlnPheAlaArgGluAsnProThrArgGluValLeuProLysVal 80
             ||||||||||||:::  ||||||:::|||||||||      ||||||||||||||||||
Db       311 CTACTCATGCGCATTCACTTTGCAAAAGAGAACCCAATGAACGAAGTCCTACCCAAAGTC 370

Qy        81 GlyValLysAspIleGluAspValThrGlnGluIleValThrLysThrLeuArgArgAla 100
             ||||||||||||||||||||||:::|||  |||      ||||||||||||||||||
Db       371 AGAGTCAAAGATATTGAAGATGTAACTGAAGAGACCGTGAAAACTACATTAAGAAGGGCC 430

Qy       101 ValSerPheHisSerThrLeuGlnCysHisAspGlyHisTrpProGlyAspTyrGlyGly 120
             ::::::|||||||||||||||||||  ||||||||||||||||||||||||||||||||
Db       431 ATAAATTTCCATTCAACTCTCCAGAGTCATGATGGACACTGGCCAGGAGATTATGGAGGT 490

Qy       121 ProMetPheLeuMetProGlyLeuValIleThrLeuSerIleThrGlyAlaLeuAsnThr 140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       491 CCCATGTTTCTTATGCCTGGCTTGGTAATTACTCTGTCTATCACCGGGGCTCTGAATGCA 550

Qy       141 ValLeuThrGluGluHisArgLysGluIleCysArgTyrLeuTyrAsnHisGlnAsnLys 160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       551 GTCTTAACAGAAGAGCATAGAAAGGAAATATGCCGTTACCTCTATAACCATCAGAACAAG 610

Qy       161 AspGlyGlyTrpGlyLeuHisIleGluGlyProSerThrMetPheGlySerValLeuSer 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db       611 GATGGTGGGTGGGGTTTACATATTGAAGGTCAAGCACCATGTTTGGCTCTGTCTTGAAT 670

Qy       181 TyrIleThrLeuArgLeuLeuGlyGluGlyProAsnAspGlyGlnGlyGluMetGluLys 200
             |||:::  ||||||||||||||||||||||||||||||  ||||||||||||||||||||
Db       671 TATGTTGCTCTGAGATTGCTTGGCGAGGGACCTAATGATAGACAAGGGGAAATGGAGAAG 730

Qy       201 AlaArgAspTrpIleLeuGlyHisGlyGlyAlaThrTyrIleThrSerTrpGlyLysMet 220
             ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db       731 GGACGTGACTGGATTCTTGGGCATGGCGGTGCTACTTTTATAACATCGTGGGGAAAGATG 790

Qy       221 TrpLeuSerValLeuGlyValTyrGluTrpSerGlyAsnAsnProLeuProProGluIle 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       791 TGGCTTTCAGTACTTGGAGTGTATGAATGGTCTGGAAATAATCCCCTGCCTCCTGAGATA 850

Qy       241 TrpLeuLeuProTyrMetLeuProPheHisProGlyArgMetTrpCysHisCysArgMet 260
             ||||||||||||||:::||||||  ||||||||||||||||||||||||||||||||||
Db       851 TGGCTCCTTCCATACGTTCTTCCAATTCATCCAGGAAGGATGTGGTGTCACTGCCGGATG 910

Qy       261 ValTyrLeuProMetSerTyrLeuTyrGlyLysArgPheValGlyProIleSerProThr 280
             ||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db       911 GTCTATTTGCCAATGTCCTACTTATATGGAAAGAGGTTTGTTGGCCCAATCACACCAACA 970

Qy       281 ValLeuSerLeuArgLysGluLeuTyrThrValProTyrHisAspIleAspTrpAspGln 300
             :::||||||||||||||||||||||||||||:::|||||||||||||||||||:::|||
Db       971 ATATTATCTTTGAGAAAGGAGCTTTATACCATACCATATCATGATATAGACTGGAATCAG 1030

Qy       301 AlaArgAsnLeuCysAlaLysGluAspLeuTyrTyrProHisProLeuValGlnAspIle 320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1031 GCTCGCAATTTGTGTGCAAAGGAAGACTTATACTATCCTCACCCACTTGTGCAAGATATT 1090

Qy       321 LeuTrpAlaSerLeuHisLysPheLeuGluProIleLeuMetHisTrpProGlyLysArg 340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db      1091 CTGTGGGCATCTCTACACAAGTTCTTGGAGCCTATTTTGATGCATTGGCCTGGCAAAAAG 1150
```

US-09-419-679-6 (1-757) x AR407878 (1-2560)

```
Qy          1 MetTrpLysLeuLysPheAlaGluGlyGlyAsnProTrpLeuArgThrLeuAsnAsnHis 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         25 ATGTGGAAGCTCAAGTTCGCCGAAGGAGGGAATCCATGGCTTCGGACATTGAACAATCAC 84

Qy         21 ValGlyArgGlnValTrpGluPheAspProLysLeuGlySerProGlnAspLeuLeuGlu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         85 GTTGGAAGACAGGTGTGGGAGTTCGATCCTAAGCTTGGATCGCCGCAAGATCTCCTCGAG 144

Qy         41 IleGluLysAlaArgGlnAsnPheHisAspAsnArgPheThrHisLysHisSerAlaAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        145 ATTGAGAAAGCTCGCCAGAATTTTCACGATAACCGCTTTACCCACAAACACAGCGCTGAT 204

Qy         61 LeuLeuMetArgMetGlnPheAlaArgGluAsnProThrArgGluValLeuProLysVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        205 CTACTTATGCGGATGCAGTTCGCAAGAGAGAACCCAACACGTGAAGTCTTGCCCAAAGTC 264

Qy         81 GlyValLysAspIleGluAspValThrGlnGluIleValThrLysThrLeuArgArgAla 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        265 GGAGTTAAGGATATTGAGGATGTGACCCAAGAGATTGTGACAAAAACATTAAGAAGGGCC 324

Qy        101 ValSerPheHisSerThrLeuGlnCysHisAspGlyHisTrpProGlyAspTyrGlyGly 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        325 GTAAGTTTCCATTCAACTCTCCAGTGCCATGACGGACACTGGCCGGGAGATTATGGAGGT 384

Qy        121 ProMetPheLeuMetProGlyLeuValIleThrLeuSerIleThrGlyAlaLeuAsnThr 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        385 CCCATGTTTCTGATGCCTGGCTTGGTAATTACTCTGTCTATCACTGGGGCGTTGAATACA 444

Qy        141 ValLeuThrGluGluHisArgLysGluIleCysArgTyrLeuTyrAsnHisGlnAsnLys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        445 GTCTTAACTGAAGAACATAGAAAGGAAATATGCCGTTACCTCTATAATCATCAAAACAAG 504

Qy        161 AspGlyGlyTrpGlyLeuHisIleGluGlyProSerThrMetPheGlySerValLeuSer 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        505 GATGGTGGGTGGGGTTTGCATATTGAAGGTCCAAGCACCATGTTTGGCTCTGTCTTGAGT 564

Qy        181 TyrIleThrLeuArgLeuLeuGlyGluGlyProAsnAspGlyGlnGlyGluMetGluLys 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        565 TATATTACTCTGAGATTGCTAGGTGAGGGGCCTAATGATGGACAAGGGGAAATGGAGAAG 624

Qy        201 AlaArgAspTrpIleLeuGlyHisGlyGlyAlaThrTyrIleThrSerTrpGlyLysMet 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        625 GCACGTGACTGGATTCTAGGGCATGGTGGTGCTACTTATATAACGTCATGGGGGAAGATG 684

Qy        221 TrpLeuSerValLeuGlyValTyrGluTrpSerGlyAsnAsnProLeuProProGluIle 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        685 TGGCTTTCAGTACTTGGAGTGTATGAATGGTCTGGAAATAATCCCCTGCCCCCTGAGATA 744

Qy        241 TrpLeuLeuProTyrMetLeuProPheHisProGlyArgMetTrpCysHisCysArgMet 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        745 TGGCTCCTTCCATACATGCTTCCATTTCATCCAGGAAGGATGTGGTGTCACTGCCGGATG 804

Qy        261 ValTyrLeuProMetSerTyrLeuTyrGlyLysArgPheValGlyProIleSerProThr 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        805 GTCTATTTGCCGATGTCCTACTTATATGGCAAGAGGTTTGTTGGTCCAATCTCACCAACA 864

Qy        281 ValLeuSerLeuArgLysGluLeuTyrThrValProTyrHisAspIleAspTrpAspGln 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        865 GTATTATCTTTGAGAAAAGAGCTTTATACAGTACCATACCATGATATAGATTGGGATCAG 924

Qy        301 AlaArgAsnLeuCysAlaLysGluAspLeuTyrTyrProHisProLeuValGlnAspIle 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        925 GCTCGCAATTTGTGTGCAAAGGAAGATTTGTACTATCCTCACCCACTTGTACAGGATATT 984

Qy        321 LeuTrpAlaSerLeuHisLysPheLeuGluProIleLeuMetHisTrpProGlyLysArg 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        985 CTTTGGGCATCTCTACACAAGTTCCTTGAGCCTATTCTGATGCATTGGCCTGGAAAAAGA 1044

Qy        341 LeuArgGluLysAlaIleIleSerAlaLeuGluHisIleHisTyrGluAspGluAsnThr 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1045 TTGAGGGAAAAGGCTATTATTTCTGCATTGGAGCATATACATTACGAAGATGAGAATACT 1104

Qy        361 ArgTyrIleCysIleGlyProValAsnLysValLeuAsnMetLeuCysCysTrpValGlu 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1105 CGATATATTTGCATAGGTCCTGTAAATAAGGTGTTAAATATGCTTTGCTGTTGGGTGGAA 1164

Qy        381 AspProAsnSerGluAlaPheLysLeuHisLeuProArgIleTyrAspTyrLeuTrpIle 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1165 GATCCAAATTCTGAGGCCTTCAAGTTGCATCTTCCCAGGATTTATGATTATCTATGGATT 1224

Qy        401 AlaGluAspGlyMetLysMetGlnGlyTyrAsnGlySerGlnLeuTrpAspThrAlaPhe 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1225 GCAGAAGATGGCATGAAAATGCAGGGCTACAATGGAAGTCAACTATGGGACACTGCTTTT 1284

Qy        421 AlaValGlnAlaIleIleAlaSerAsnLeuIleGluGluPheGlyProThrIleArgLys 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1285 GCTGTCCAAGCAATTATTGCATCTAACCTCATTGAAGAATTTGGTCCAACTATAAGAAAA 1344

Qy        441 AlaHisThrTyrIleLysAsnSerGlnValLeuGluAspCysProGlyAspLeuAsnLys 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1345 GCTCATACCTATATTAAGAATTCACAGGTTTTAGAAGATTGTCCAGGTGATCTTAATAAA 1404

Qy        461 TrpTyrArgHisIleSerLysGlyAlaTrpProPheSerThrGlyAspHisGlyTrpPro 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1405 TGGTACCGTCACATTTCAAAAGGTGCTTGGCCTTTTTCAACTGGAGATCATGGATGGCCA 1464

Qy        481 IleSerAspCysThrAlaGluGlyLeuLysAlaValLeuLeuLeuSerLysIleAlaPro 500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1465 ATTTCTGACTGCACAGCTGAAGGACTGAAAGCTGTTCTATTACTATCCAAAATTGCACCA 1524

Qy        501 GluIleValGlyGluProIleAspValLysArgLeuTyrAspSerValAsnValIleLeu 520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1525 GAAATAGTTGGTGAGCCAATAGACGTGAAGCGATTATATGATTCTGTAAATGTCATTCTC 1584

Qy        521 SerLeuGlnAsnGluAspGlyGlyPheAlaThrTyrGluLeuLysArgSerTyrAsnTrp 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1585 TCACTACAGAATGAAGATGGTGGTTTTGCAACATATGAGCTTAAACGATCTTATAATTGG 1644

Qy        541 LeuGluIleIleAsnProAlaGluThrPheGlyAspIleValIleAspTyrProTyrVal 560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1645 TTGGAGATAATCAATCCTGCTGAAACTTTTGGTGACATCGTTATTGATTATCCTTATGTG 1704

Qy        561 GluCysThrSerAlaAlaIleGlnAlaLeuAlaSerPheArgLysLeuTyrProGlyHis 580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1705 GAATGTACATCAGCAGCGATTCAAGCTTTGGCATCATTTAGGAAATTATATCCTGGGCAT 1764

Qy        581 ArgArgGluGluIleGlnHisCysIleAspLysAlaThrThrPheIleGluLysIleGln 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1765 CGCCGAGAAGAAATACAACATTGTATCGATAAAGCCACTACCTTCATTGAAAAAATACAA 1824

Qy        601 AlaSerAspGlySerTrpTyrGlySerTrpGlyValCysPheThrTyrGlyAlaTrpPhe 620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1825 GCTTCAGATGGATCATGGTATGGTTCTTGGGGAGTTTGCTTCACTTACGGTGCTTGGTTT 1884

Qy        621 GlyValLysGlyLeuIleAlaAlaGlyArgSerPheSerAsnCysSerSerIleArgLys 640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1885 GGGGTAAAAGGTCTGATTGCTGCTGGAAGGAGTTTCAGTAATTGCTCAAGCATCCGTAAA 1944

Qy        641 AlaCysGluPheLeuLeuSerLysGlnLeuProSerGlyGlyTrpGlyGluSerTyrLeu 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1945 GCTTGTGAATTTCTGCTGTCCAAGCAGCTTCCTTCTGGTGGCTGGGGAGAGAGTTATCTG 2004

Qy        661 SerCysGlnAsnLysValTyrSerAsnLeuGluGlyAsnArgSerHisValValAsnThr 680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2005 TCCTGTCAAAACAAGGTGTATTCAAATCTGGAAGGCAACAGGTCTCATGTGGTCAACACT 2064

Qy        681 GlyTrpAlaMetLeuAlaLeuIleAspAlaGlyGlnAlaLysArgAspSerGlnProLeu 700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2065 GGGTGGGCTATGTTGGCTCTCATTGATGCTGGACAGGCTAAGAGAGATTCGCAACCACTG 2124

Qy        701 HisArgAlaAlaAlaTyrLeuIleAsnSerGlnLeuGluAspGlyAspPheProGlnGln 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2125 CACCGGGCAGCTGCATACTTGATAAATTCCCAATTGGAGGACGGTGACTTTCCGCAGCAG 2184
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 5, 2005, 09:55:37 ; Search time 8520 Seconds
(without alignments)
4305.235 Million cell updates/sec

Title: US-09-419-679-6
Perfect score: 4148
Sequence: 1 MWKLKFAEGGNPWLRTLNNH.....RNIFPIWALGEYQSQVLQSR 757

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/6630617/runat_02112005_120300_15076/app_query.fasta_1.903
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=6630617_@CGN_1_1_4892_@runat_02112005_120300_15076 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

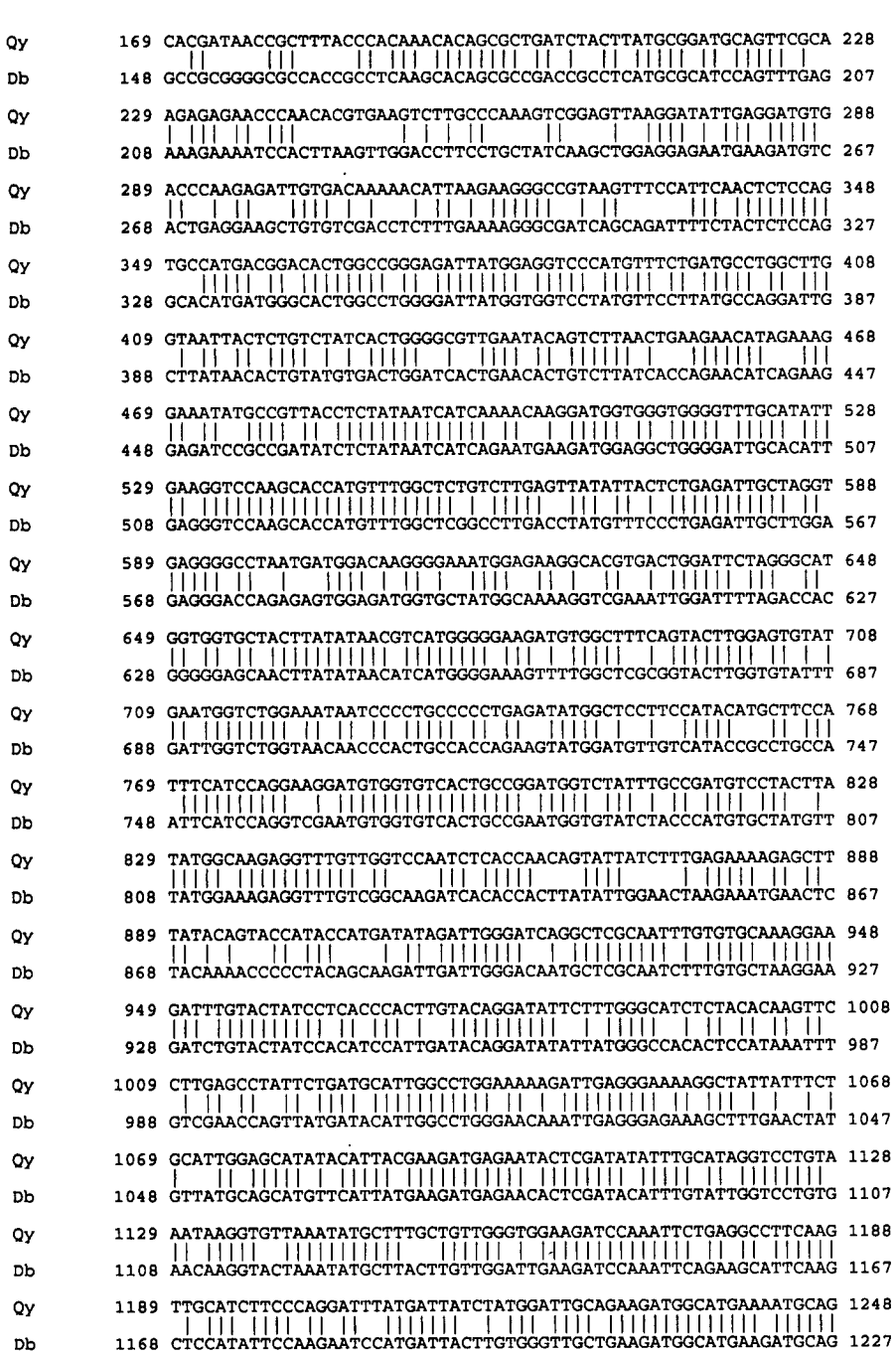
Result No.	Score	% Query Match	Length	DB	ID	Description
1	4148	100.0	2560	6	AR407878	AR407878 Sequence
2	3837	92.5	2889	8	AB025968	AB025968 Glycyrrhi
3	3678.5	88.7	2677	8	D89619	D89619 Pisum sativ
4	3537	85.3	2547	8	AY520819	AY520819 Centella
5	3528	85.1	2559	8	AB009029	AB009029 Panax gin
6	3518	84.8	2514	8	AB055510	AB055510 Betula pl
7	3486	84.0	2561	8	U02555	U02555 Arabidopsis
8	3480	83.9	2280	6	AX506138	AX506138 Sequence
9	3480	83.9	2280	8	BT001118	BT001118 Arabidops
10	3480	83.9	2654	8	AY094394	AY094394 Arabidops
11	3373	81.3	2556	8	AB116237	AB116237 Cucurbita
12	3352	80.8	2683	8	AB058507	AB058507 Costus sp
13	3331	80.3	2433	8	AB033334	AB033334 Luffa cyl
14	3323	80.1	2460	8	AB055509	AB055509 Betula pl
15	3265.5	78.7	2280	8	AY618692	AY618692 Avena pro
16	3260.5	78.6	2280	8	AY618691	AY618691 Avena lon
17	3260.5	78.6	2280	8	AY618693	AY618693 Avena str
18	3260.5	78.6	2280	8	AST311790	AJ311790 Avena str
19	3256.5	78.5	2651	8	AK121211	AK121211 Oryza sat
20	3254.5	78.5	2280	8	AY618690	AY618690 Avena lon
21	3240.5	78.1	2413	8	AF169966	AF169966 Oryza sat
22	3232.5	77.9	2280	8	AY618689	AY618689 Avena cla
23	3231.5	77.9	2280	8	AY618694	AY618694 Avena ven
24	3184	76.8	3057	8	AF216755	AF216755 Abies mag
25	3125	75.3	2558	6	AR407876	AR407876 Sequence
26	3122	75.3	2579	8	AB025353	AB025353 Allium ma
27	3004.5	72.4	2487	8	AB058508	AB058508 Costus sp
28	2979	71.8	2444	8	AB116238	AB116238 Cucurbita
29	2927.5	70.6	2684	8	AB009031	AB009031 Panax gin
30	2846	68.6	2347	8	AB033335	AB033335 Luffa cyl
31	2787	67.2	2578	8	AK066327	AK066327 Oryza sat
32	2783	67.1	1983	8	AB025344	AB025344 Olea euro
33	2767	66.7	2458	8	AK103080	AK103080 Oryza sat
34	2751	66.3	2493	8	AB025346	AB025346 Taraxacum
35	2731.5	65.9	2409	8	AB116239	AB116239 Cucurbita
36	2718	65.5	2546	8	AB025343	AB025343 Olea euro
37	2646.5	63.8	2558	8	AK068026	AK068026 Oryza sat
38	2611.5	63.0	2452	8	AB025345	AB025345 Taraxacum
39	2598.5	62.6	2471	8	AB055511	AB055511 Betula pl
40	2597	62.6	2519	8	AB055512	AB055512 Betula pl
41	2566.5	61.9	2589	8	AB009030	AB009030 Panax gin
42	2551	61.5	2538	6	AX382063	AX382063 Sequence
43	2543.5	61.3	2657	8	AB116228	AB116228 Glycyrrhi
44	2538	61.2	2478	6	AX382059	AX382059 Sequence
45	2519.5	60.7	2729	8	AB014057	AB014057 Panax gin

ALIGNMENTS

RESULT 1
AR407878
LOCUS AR407878 2560 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 5 from patent US 6630617.
ACCESSION AR407878
VERSION AR407878.1 GI:40157830
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 2560)
AUTHORS Famodu,O.O., Kinney,A.J. and Rafalski,J.A.
TITLE Enzymes involved in squalene metabolism
JOURNAL Patent: US 6630617-A 5 07-OCT-2003;
FEATURES Location/Qualifiers
source 1. .2560
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.: 0 Length: 2560
Score: 4148.00 Matches: 757
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

Search completed: November 5, 2005, 09:54:37
Job time : 10835 secs


Qy 169 CACGATAACCGCTTTACCCACAAACACAGCGCTGATCTACTTATGCGGATGCAGTTCGCA 228
Db 148 GCCGCGGGGCGCCACCGCCTCAAGCACAGCGCCGACCGCCTCATGCGCATCCAGTTTGAG 207
Qy 229 AGAGAGAACCCAACACGTGAAGTCTTGCCCAAAGTCGGAGTTAAGGATATTGAGGATGTG 288
Db 208 AAAGAAAATCCACTTAAGTTGGACCTTCCTGCTATCAAGCTGGAGGAGAATGAAGATGTC 267
Qy 289 ACCCAAGAGATTGTGACAAAAACATTAAGAAGGGCCGTAAGTTTCCATTCAACTCTCCAG 348
Db 268 ACTGAGGAAGCTGTGTCGACCTCTTTGAAAAGGGCGATCAGCAGATTTTCTACTCTCCAG 327
Qy 349 TGCCATGACGGACACTGGCCGGGAGATTATGGAGGTCCCATGTTTCTGATGCCTGGCTTG 408
Db 328 GCACATGATGGGCACTGGCCTGGGGATTATGGTGGTCCTATGTTCCTTATGCCAGGATTG 387
Qy 409 GTAATTACTCTGTCTATCACTGGGGCGTTGAATACAGTCTTAACTGAAGAACATAGAAAG 468
Db 388 CTTATAACACTGTATGTGACTGGATCACTGAACACTGTCTTATCACCAGAACATCAGAAG 447
Qy 469 GAAATATGCCGTTACCTCTATAATCATCAAAACAAGGATGGTGGGTGGGGTTTGCATATT 528
Db 448 GAGATCCGCCGATATCTCTATAATCATCAGAATGAAGATGGAGGCTGGGGATTGCACATT 507
Qy 529 GAAGGTCCAAGCACCATGTTTGGCTCTGTCTTGAGTTATATTACTCTGAGATTGCTAGGT 588
Db 508 GAGGGTCCAAGCACCATGTTTGGCTCGGCCTTGACCTATGTTTCCCTGAGATTGCTTGGA 567
Qy 589 GAGGGGCCTAATGATGGACAAGGGGAAATGGAGAAGGCACGTGACTGGATTCTAGGGCAT 648
Db 568 GAGGGACCAGAGAGTGGAGATGGTGCTATGGCAAAAGGTCGAAATTGGATTTTAGACCAC 627
Qy 649 GGTGGTGCTACTTATATAACGTCATGGGGGAAGATGTGGCTTTCAGTACTTGGAGTGTAT 708
Db 628 GGGGGAGCAACTTATATAACATCATGGGGAAAGTTTTGGCTCGCGGTACTTGGTGTATTT 687
Qy 709 GAATGGTCTGGAAATAATCCCCTGCCCCCTGAGATATGGCTCCTTCCATACATGCTTCCA 768
Db 688 GATTGGTCTGGTAACAACCCACTGCCACCAGAAGTATGGATGTTGTCATACCGCCTGCCA 747
Qy 769 TTTCATCCAGGAAGGATGTGGTGTCACTGCCGGATGGTCTATTTGCCGATGTCCTACTTA 828
Db 748 ATTCATCCAGGTCGAATGTGGTGTCACTGCCGAATGGTGTATCTACCCATGTGCTATGTT 807
Qy 829 TATGGCAAGAGGTTTGTTGGTCCAATCTCACCAACAGTATTATCTTTGAGAAAAGAGCTT 888
Db 808 TATGGAAAGAGGTTTGTCGGCAAGATCACACCACTTATATTGGAACTAAGAAATGAACTC 867
Qy 889 TATACAGTACCATACCATGATATAGATTGGGATCAGGCTCGCAATTTGTGTGCAAAGGAA 948
Db 868 TACAAAACCCCCTACAGCAAGATTGATTGGGACAATGCTCGCAATCTTTGTGCTAAGGAA 927
Qy 949 GATTTGTACTATCCTCACCCACTTGTACAGGATATTCTTTGGGCATCTCTACACAAGTTC 1008
Db 928 GATCTGTACTATCCACATCCATTGATACAGGATATATTATGGGCCACACTCCATAAATTT 987
Qy 1009 CTTGAGCCTATTCTGATGCATTGGCCTGGAAAAAGATTGAGGGAAAAGGCTATTATTTCT 1068
Db 988 GTCGAACCAGTTATGATACATTGGCCTGGGAACAAATTGAGGGAGAAAGCTTTGAACTAT 1047
Qy 1069 GCATTGGAGCATATACATTACGAAGATGAGAATACTCGATATATTTGCATAGGTCCTGTA 1128
Db 1048 GTTATGCAGCATGTTCATTATGAAGATGAGAACACTCGATACATTTGTATTGGTCCTGTG 1107
Qy 1129 AATAAGGTGTTAAATATGCTTTGCTGTTGGGTGGAAGATCCAAATTCTGAGGCCTTCAAG 1188
Db 1108 AACAAGGTACTAAATATGCTTACTTGTTGGATTGAAGATCCAAATTCAGAAGCATTCAAG 1167
Qy 1189 TTGCATCTTCCCAGGATTTATGATTATCTATGGATTGCAGAAGATGGCATGAAAATGCAG 1248
Db 1168 CTCCATATTCCAAGAATCCATGATTACTTGTGGGTTGCTGAAGATGGCATGAAGATGCAG 1227
Qy 1249 GGCTACAATGGAAGTCAACTATGGGACACTGCTTTTGCTGTCCAAGCAATTATTGCATCT 1308
Db 1228 GGTTATAATGGCAGCCAGCTCTGGGACACAGCTTTCGCAGTCCAAGCTATTGCGGCTACT 1287
Qy 1309 AACCTCATTGAAGAATTTGGTCCAACTATAAGAAAAGCTCATACCTATATTAAGAATTCA 1368
Db 1288 GGCCTCATTGATGAGTTTGCTCCCACTCTTAAGCTAGCACACAACTTCATTAAGAACTCC 1347
Qy 1369 CAGGTTTTAGAAGATTGTCCAGGTGATCTTAATAAATGGTACCGTCACATTTCAAAAGGT 1428
Db 1348 CAGGTTCTTGATGACTGCCCCGGAGATCTGAGTTACTGGTACCGTCACATATCTAAAGGT 1407
Qy 1429 GCTTGGCCTTTTTCAACTGGAGATCATGGATGGCCAATTTCTGACTGCACAGCTGAAGGA 1488
Db 1408 GCATGGCCTTTTTCTACTGCTGATCATGGTTGGCCTATATCAGATTGCACTGCAGAAGGA 1467
Qy 1489 CTGAAAGCTGTTCTATTACTATCCAAAATTGCACCAGAAATAGTTGGTGAGCCAATAGAC 1548
Db 1468 CTAAAGGCTGCATTATTGCTATCGAAGATTTCTCCAGAAATTGTGGGCGAACCAGTGGAA 1527
Qy 1549 GTGAAGCGATTATATGATTCTGTAAATGTCATTCTCTCACTACAGAATGAAGATGGTGGT 1608
Db 1528 GTTAACAGACTTTATGATGCTGTCAATTGTCTGATGTCTTGGATGAACAATAATGGTGGT 1587
Qy 1609 TTTGCAACATATGAGCTTAAACGATCTTATAATTGGTTGGAGATAATCAATCCTGCTGAA 1668
Db 1588 TTTGCAACATACGAACTCACAAGGTCTTATGCCTGGCTGGAGCTTATCAACCCTGCTGAG 1647
Qy 1669 ACTTTTGGTGACATCGTTATTGATTATCCTTATGTGGAATGTACATCAGCAGCGATTCAA 1728
Db 1648 ACCTTTGGAGACATTGTCATTGATTACCCGTATGTGGAATGTACTTCAGCAGCAATTCAG 1707
Qy 1729 GCTTTGGCATCATTTAGGAAATTATATCCTGGGCATCGCCGAGAAGAAATACAACATTGT 1788
Db 1708 GCCCTGACATCATTTAAAAAGCTCTATCCTGGGCACCGCAGGAAAGATGTAGATAACTGT 1767
Qy 1789 ATCGATAAAGCCACTACCTTCATTGAAAAAATACAAGCTTCAGATGGATCATGGTATGGT 1848
Db 1768 ATCAACAAAGCTGCTAATTACATAGAGAGTATTCAAAGAAGTGATGGTTCATGGTATGGC 1827
Qy 1849 TCTTGGGGAGTTTGCTTCACTTACGGTGCTTGGTTTGGGGTAAAAGGTCTGATTGCTGCT 1908
Db 1828 TCTTGGGCTGTTTGTTTCACCTACGCCACATGGTTTGGAGTGAAGGCATTAGTTGCTGCT 1887
Qy 1909 GGAAGGAGTTTCAGTAATTGCTCAAGCATCCGTAAAGCTTGTGAATTTCTGCTGTCCAAG 1968
Db 1888 GGTAGGCCGTTCAAGAGTAGTCCTGCAATCAGGAAGGCATGTGAATTTCTGATGTCAAAA 1947
Qy 1969 CAGCTTCCTTCTGGTGGCTGGGGAGAGAGTTATCTGTCCTGTCAAAACAAGGTGTATTCA 2028
Db 1948 GAGCTTCCTTTCGGTGGCTGGGGAGAAAGTTATTTGTCATGTCGAGATCAGGTTTACACT 2007
Qy 2029 AATCTGGAAGGCAACAGGTCTCATGTGGTCAACACTGGGTGGGCTATGTTGGCTCTCATT 2088
Db 2008 AATCTTGAAGGCAAGCATGCTCACGCAGTCAACACTGGCTGGGCCATGCTGACCCTAATT 2067
Qy 2089 GATGCTGGACAGGCTAAGAGAGATTCGCAACCACTGCACCGGGCAGCTGCATACTTGATA 2148
Db 2068 GACGCAGGACAGGCTGAGAGAGATCCAGCACCTCTGCATCGAGCAGCGAAGGTTTTGATC 2127
Qy 2149 AATTCCCAATTGGAGGACGGTGACTTTCCGCAGCAGGAAATAATGGGAGTCTTCAACAAG 2208
Db 2128 AACTTACAATCTGAGGATGGGGAATTTCCTCAACAAGAGATCATGGGAGTCTTCAACAAA 2187
Qy 2209 AATTGCATGATCACATACGCCGCATACAGAAACATATTCCCCATTTGGGCGTTGGGAGAA 2268
Db 2188 AACTGCATGATCAGTTACTCCCAGTACAGGGACATCTTCCCTGTGTGGGCACTCGGAGAG 2247
Qy 2269 TACCAATCCCAAGTATTG 2286
Db 2248 TACCGTTGCCGAGTCTTG 2265

```
Qy       1103 CTCGATATATTTGCATAGGTCCTGTAAATAAGGTGTTAAATATGCTTTGCTGTTGGGTGG 1162
Db       1285 CAAGATATATTTGCATTGGTCCTGTCAACAAGGTACTTAATATGCTTTGCTGCTGGATTG 1344

Qy       1163 AAGATCCAAATTCTGAGGCCTTCAAGTTGCATCTTCCCAGGATTTATGATTATCTATGGA 1222
Db       1345 AAGATCCAAATTCAGAGGCATTCAAGCTACATCTTCCTAGAGTAATGGATTATCTGTGGT 1404

Qy       1223 TTGCAGAAGATGGCATGAAAATGCAGGGCTACAATGGAAGTCAACTATGGGACACTGCTT 1282
Db       1405 TGGCTGAAGATGGAATGAAAATGCGGGGCTATAATGGCAGCCAGTTGTGGGACACAGCTT 1464

Qy       1283 TTGCTGTCCAAGCAATTATTGCATCTAACCTCATTGAAGAATTTGGTCCAACTATAAGAA 1342
Db       1465 TCACTGTTCAGGCAATAATATCGACTGACCTTTTTGAGGAATTTGGTCTTGCACTGACGA 1524

Qy       1343 AAGCTCATACCTATATTAAGAATTCACAGGTTTTAGAAGATTGTCCAGGTGATCTTAATA 1402
Db       1525 AGGCACATGAATTTATTAAAAAAACACAGGTTCTTGAAGACTGCCCTGGCGATCTTAATT 1584

Qy       1403 AATGGTACCGTCACATTTCAAAAGGTGCTTGGCCTTTTTCAACTGGAGATCATGGATGGC 1462
Db       1585 TTTGGTACCGCCACATCTCTAAAGGTGCATGGCCTTTCTCCACCGCTGATCATGGATGGC 1644

Qy       1463 CAATTTCTGACTGCACAGCTGAAGGACTGAAAGCTGTTCTATTACTATCCAAAATTGCAC 1522
Db       1645 CTATCTCAGATTGCACTTCCGAAGGACTTAAGGCAGCGCTCTTATTATCTAAGATTTCTC 1704

Qy       1523 CAGAAATAGTTGGTGAGCCAATAGACGTGAAGCGATTATATGATTCTGTAAATGTCATTC 1582
Db       1705 CTGAAATTGTTGGTGATCCACTTGATGGAAAGAGCCTCTATGATGCCGTTAATGTTATCC 1764

Qy       1583 TCTCACTACAGAATGAAGATGGTGGTTTTGCAACATATGAGCTTAAACGATCTTATAATT 1642
Db       1765 TTTCACTAATGAATAATGATGGTGGTTTTGCTACGTATGAACTTACGAGATCCTATGCCT 1824

Qy       1643 GGTTGGAGATAATCAATCCTGCTGAAACTTTTGGTGACATCGTTATTGATTATCCTTATG 1702
Db       1825 GGTTGGAGATTATCAATCCTGCTGAAACATTTGGAGACATTGTTATCGATTACCCATATG 1884

Qy       1703 TGGAATGTACATCAGCAGCGATTCAAGCTTTGGCATCATTTAGGAAATTATATCCTGGGC 1762
Db       1885 TTGAGTGCACTTCGGCAGCAATTCAGGCATTGACATTATTTAAAAAGACATATCCTGGAC 1944

Qy       1763 ATCGCCGAGAAGAAATACAACATTGTATCGATAAAGCCACTACCTTCATTGAAAAAATAC 1822
Db       1945 ATCGCCGAGAGGAGATAGACAATTGCATCAGAAAATCAGCTCGTTTTATTGAAAAGATTC 2004

Qy       1823 AAGCTTCAGATGGATCATGGTATGGTTCTTGGGGAGTTTGCTTCACTTACGGTGCTTGGT 1882
Db       2005 AGCTAGCTGATGGTTCATGGTATGGATCTTGGGGTGTTTGCTTCACATATGGTATATGGT 2064

Qy       1883 TTGGGGTAAAAGGTCTGATTGCTGCTGGAAGGAGTTTCAGTAATTGCTCAAGCATCCGTA 1942
Db       2065 TTGGAATGAAGGGATTACTTGCCGCTGGTCGGACCTATGAGACTAGCTCTTGCATCCGAA 2124

Qy       1943 AAGCTTGTGAATTTCTGCTGTCCAAGCAGCTTCCTTCTGGTGGCTGGGGAGAGAGTTATC 2002
Db       2125 AGGCATGTGACTTTCTGTTGTCTAAACAAGTTGCCTCTGGTGGTTGGGGGGAGAGCTATC 2184

Qy       2003 TGTCCTGTCAAAACAAGGTGTATTCAAATCTGGAAGGCAACAGGTCTCATGTGGTCAACA 2062
Db       2185 TTTCATGTCAAAATAAGGTCTACACAAATCTTGAGGGTAACAGGGCTCATGCAGTGAATA 2244

Qy       2063 CTGGGTGGGCTATGTTGGCTCTCATTGATGCTGGACAGGCTAAGAGAGATTCGCAACCAC 2122
Db       2245 CTGGCTGGGCCATGCTGGCTCTAATTGATGCTGGGCAGGGTGAGAGAGATCCAAAACCAT 2304

Qy       2123 TGCACCGGGCAGCTGCATACTTGATAAATTCCCAATTGGAGGACGGTGACTTTCCGCAGC 2182
Db       2305 TGCACAGAGCAGCAAAAGTTCTAATAAACATGCAAATGGAGAATGGCGAGTTTCCCCAAC 2364
```

```
Qy       2183 AGGAAATAATGGGAGTCTTCAACAAGAATTGCATGATCACATACGCCGCATACAGAAACA 2242
Db       2365 AAGAGATTATGGGGGTTTTTAACAAAAATTGCATGATCAGTTATTCAGAATATCGAAACA 2424

Qy       2243 TATTCCCCATTTGGGCGTTGGGAGAATACCAATCCCAAGTATTGCAATCTC 2293
Db       2425 TATTTCCAATCTGGGCTCTCGGAGAATACCGAAGACGCGTCCTGTGTTCTC 2475
```

```
RESULT 15
AY618689
LOCUS       AY618689               2280 bp    DNA     linear   PLN 06-JUN-2004
DEFINITION  Avena clauda strain CAV5566/1 cycloartenol synthase gene, complete
            cds.
ACCESSION   AY618689
VERSION     AY618689.1  GI:47834378
KEYWORDS    .
SOURCE      Avena clauda
  ORGANISM  Avena clauda
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Aveneae; Avena.
REFERENCE   1  (bases 1 to 2280)
  AUTHORS   Qi,X., Bakht,S., Leggett,M., Maxwell,C., Melton,R. and Osbourn,A.
  TITLE     A gene cluster for secondary metabolism in oat: Implications for
            the evolution of metabolic diversity in plants
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (21), 8233-8238 (2004)
   PUBMED   15148404
REFERENCE   2  (bases 1 to 2280)
  AUTHORS   Qi,X., Bakht,S., Leggett,M., Maxwell,C., Melton,R. and Osbourn,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAY-2004) Sainsbury Laboratory, John Innes Centre,
            Norwich, Norfolk NR4 7UH, UK
FEATURES             Location/Qualifiers
     source          1. .2280
                     /organism="Avena clauda"
                     /mol_type="genomic DNA"
                     /strain="CAV5566/1"
                     /db_xref="taxon:83523"
     mRNA            <1. .>2280
                     /product="cycloartenol synthase"
     CDS             1. .2280
                     /codon_start=1
                     /product="cycloartenol synthase"
                     /protein_id="AAT38887.1"
                     /db_xref="GI:47834379"
                     /translation="MWRLKIAEGGGDPWLRTKNAHVGRQVWEFDPEAGDPEALAAVEA
                     ARREFAAGRHRLKHSADRLMRIQFEKENPLKLDLPAIKLEENEDVTEEAVSTSLKRAI
                     SRFSTLQAHDGHWPGDYGGPMFLMPGLLITLYVTGSLNTVLSPEHQKEIRRYLYNHQN
                     EDGGWGLHIEGPSTMFGSALTYVSLRLLGEGPESGDGAMAKGRNWILDHGGATYITSW
                     GKFWLAVLGVFDWSGNNPLPPEVWMLSYRLPIHPGRMWCHCRMVYLPMCYVYGKRFVG
                     KITPLILELRNELYKTPYSKIDWDNARNLCAKEDLYYPHPLIQDILWATLHKFVEPVM
                     IHWPGNKLREKALNYVMQHVHYEDENTRYICIGPVNKVLNMLTCWIEDPNSEAFKLHI
                     PRIHDYLWVAEDGMKMQGYNGSQLWDTAFAVQAIAATGLIDEFAPTLKLAHNFIKNSQ
                     VLDDCPGDLSYWYRHISKGAWPFSTADHGWPISDCTAEGLKAALLLSKISPEIVGEPV
                     EVNRLYDAVNCLMSWMNNNGGFATYELTRSYAWLELINPAETFGDIVIDYPYVECTSA
                     AIQALTSFKKLYPGHRRKDVDNCINKAANYIESIQRSDGSWYGSWAVCFTYATWFGVK
                     ALVAAGRPFKSSPAIRKACEFLMSKELPFGGWGESYLSCRDQVYTNLEGKHAHAVNTG
                     WAMLTLIDAGQAERDPAPLHRAAKVLINLQSEDGEFPQQEIMGVFNKNCMISYSQYRD
                     IFPVWALGEYRCRVLAAGK"
ORIGIN

  Query Match             47.9%;  Score 1226.6;  DB 8;  Length 2280;
  Best Local Similarity   71.7%;  Pred. No. 1.6e-296;
  Matches 1604;  Conservative    0;  Mismatches  634;  Indels    0;  Gaps    0;

Qy         49 GGAGGGAATCCATGGCTTCGGACATTGAACAATCACGTTGGAAGACAGGTGTGGGAGTTC 108
Db         28 GGGGGCGACCCGTGGCTGCGCACGAAGAACGCGCACGTCGGCCGCCAGGTCTGGGAGTTC 87

Qy        109 GATCCTAAGCTTGGATCGCCGCAAGATCTCCTCGAGATTGAGAAAGCTCGCCAGAATTTT 168
Db         88 GACCCCGAGGCCGGCGACCCGGAGGCCCTCGCCGCCGTCGAGGCCGCGCGCCGGGAGTTC 147
```

```
Db       2222 TAGTTATGCCGCTTATCGTAACATTTTTCCTATATGGGCCCTTGGAGAATATCGCTGCCG 2281

Qy       2280 AGTATTGCAATCTCGTTAA 2298

Db       2282 GGTACTGCAAGCCCCTTAA 2300
```

RESULT 14
AB058507
LOCUS AB058507 2683 bp mRNA linear PLN 10-APR-2002
DEFINITION Costus speciosus CSOSC1 mRNA for cycloartenol synthase, complete cds.
ACCESSION AB058507
VERSION AB058507.1 GI:18147770
KEYWORDS .
SOURCE Costus speciosus
ORGANISM Costus speciosus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Zingiberales; Costaceae; Costus.
REFERENCE 1
AUTHORS Kawano,N., Ichinose,K. and Ebizuka,Y.
TITLE Molecular cloning and functional expression of cDNAs encoding oxidosqualene cyclases from Costus speciosus
JOURNAL Biol. Pharm. Bull. 25 (4), 477-482 (2002)
MEDLINE 21990915
PUBMED 11995928
REFERENCE 2 (bases 1 to 2683)
AUTHORS Kawano,N., Ichinose,K. and Ebizuka,Y.
TITLE Direct Submission
JOURNAL Submitted (23-MAR-2001) Noriaki Kawano, The University of Tokyo, Graduate School of Pharmaceutical Sciences; 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan (E-mail:ff17004@mail.ecc.u-tokyo.ac.jp, Tel:81-3-5841-4742, Fax:81-3-5841-4744)
FEATURES Location/Qualifiers
source 1. .2683
/organism="Costus speciosus"
/mol_type="mRNA"
/db_xref="taxon:49577"
/clone="CSOSC1"
/tissue_type="whole plantlet"
/dev_stage="plantlet"
/note="in vitro cultured plantlet"
gene 1. .2683
/gene="CSOSC1"
CDS 207. .2486
/gene="CSOSC1"
/EC_number="5.4.99.8"
/note="functional expression in Saccharomyces cerevisiae"
/codon_start=1
/evidence=experimental
/product="cycloartenol synthase"
/protein_id="BAB83253.1"
/db_xref="GI:18147771"
/translation="MWRLKIAEGGGPWLRTKNNHVGRQVWEFDPSLGTPEEIAEVERV
REAFRETRFEKKHSADLLMRLQFAKENPLEMNYPIIKIEEHEDVTEELVVTSLRKAIS
RVSTLQAHDGHWPGDYGGPMFLMPGLIITLYVTGALNTVLTSEHQKEIRRYLYNHQNE
DGGWGLHIEGESTMFGSALTYVILRLLGEGPDDGDGAMEKGRKWILDHGSATAITSWG
KMWLSVLGVFDWSGNNPLPPEMWLLPYFLPVHPGRMWCHCRMVYLPMSYIYGKRFTGP
ITPLILSLRKELFNLPFDQLDWNKARNECAKEDLYYPHPFIQDVLWASLHKFVEPILM
HWPGSKLREKAVNTAMQHVHYEDENTRYICIGPVNKVLNMLCCWIEDPNSEAFKLHLP
RVMDYLWLAEDGMKMRGYNGSQLWDTAFTVQAIISTDLFEEFGLALTKAHEFIKKTQV
LEDCPGDLNFWYRHISKGAWPFSTADHGWPISDCTSEGLKAALLLSKISPEIVGDPLD
GKSLYDAVNVILSLMNNDGGFATYELTRSYAWLEIINPAETFGDIVIDYPYVECTSAA
IQALTLFKKTYPGHRREEIDNCIRKSARFIEKIQLADGSWYGSWGVCFTYGIWFGMKG
LLAAGRTYETSSCIRKACDFLLSKQVASGGWGESYLSCQNKVYTNLEGNRAHAVNTGW
AMLALIDAGQGERDPKPLHRAAKVLINMQMENGEFPQQEIMGVFNKNCMISYSEYRNI
FPIWALGEYRRRVLCSQFH"
ORIGIN

Query Match 51.0%; Score 1306.2; DB 8; Length 2683;
Best Local Similarity 73.4%; Pred. No. 0;

Matches 1668; Conservative 0; Mismatches 603; Indels 0; Gaps 0;

```
Qy         23 GAATGTGGAAGCTCAAGTTCGCCGAAGGAGGGAATCCATGGCTTCGGACATTGAACAATC 82
Db        205 GAATGTGGAGGCTCAAGATCGCGGAGGGCGGCGGCCCGTGGCTGCGGACGAAGAACAACC 264

Qy         83 ACGTTGGAAGACAGGTGTGGGAGTTCGATCCTAAGCTTGGATCGCCGCAAGATCTCCTCG 142
Db        265 ACGTCGGGAGGCAGGTGTGGGAGTTCGACCCGAGCCTCGGCACGCCCGAGGAGATAGCGG 324

Qy        143 AGATTGAGAAAGCTCGCCAGAATTTTCACGATAACCGCTTTACCCACAAACACAGCGCTG 202
Db        325 AGGTGGAGAGAGTCCGGGAGGCTTTCCGCGAGACCCGGTTCGAGAAGAAGCACAGCGCCG 384

Qy        203 ATCTACTTATGCGGATGCAGTTCGCAAGAGAGAACCCAACACGTGAAGTCTTGCCCAAAG 262
Db        385 ATCTCCTCATGCGGTTGCAGTTTGCCAAGGAGAATCCACTTGAAATGAACTATCCTATCA 444

Qy        263 TCGGAGTTAAGGATATTGAGGATGTGACCCAAGAGATTGTGACAAAAACATTAAGAAGGG 322
Db        445 TCAAGATTGAAGAACACGAAGATGTCACTGAAGAACTTGTGGTGACCTCCTTAAGAAAGG 504

Qy        323 CCGTAAGTTTCCATTCAACTCTCCAGTGCCATGACGGACACTGGCCGGGAGATTATGGAG 382
Db        505 CTATCAGTCGCGTTTCTACTCTCCAAGCTCACGATGGACACTGGCCTGGGGATTATGGTG 564

Qy        383 GTCCCATGTTTCTGATGCCTGGCTTGGTAATTACTCTGTCTATCACTGGGGCGTTGAATA 442
Db        565 GTCCAATGTTCCTTATGCCGGGATTGATAATAACTTTGTATGTCACTGGAGCTCTAAACA 624

Qy        443 CAGTCTTAACTGAAGAACATAGAAAGGAAATATGCCGTTACCTCTATAATCATCAAAACA 502
Db        625 CTGTTCTTACATCTGAACACCAGAAGGAGATCCGTCGATACCTCTACAACCACCAGAATG 684

Qy        503 AGGATGGTGGGTGGGGTTTGCATATTGAAGGTCCAAGCACCATGTTTGGCTCTGTCTTGA 562
Db        685 AAGACGGAGGCTGGGGCTTGCATATTGAGGGTGAAAGCACTATGTTTGGTTCTGCCTTGA 744

Qy        563 GTTATATTACTCTGAGATTGCTAGGTGAGGGGCCTAATGATGGACAAGGGGAAATGGAGA 622
Db        745 CTTATGTTATTTTGAGATTGCTCGGAGAAGGACCTGATGATGGAGATGGGGCAATGGAAA 804

Qy        623 AGGCACGTGACTGGATTCTAGGGCATGGTGGTGCTACTTATATAACGTCATGGGGGAAGA 682
Db        805 AAGGGCGAAAGTGGATCCTAGATCATGGCAGTGCAACTGCTATAACCTCATGGGGAAAGA 864

Qy        683 TGTGGCTTTCAGTACTTGGAGTGTATGAATGGTCTGGAAATAATCCCCTGCCCCCTGAGA 742
Db        865 TGTGGTTATCTGTGCTTGGAGTGTTTGACTGGTCTGGCAACAATCCTTTGCCGCCAGAAA 924

Qy        743 TATGGCTCCTTCCATACATGCTTCCATTTCATCCAGGAAGGATGTGGTGTCACTGCCGGA 802
Db        925 TGTGGTTATTACCATACTTCTTGCCAGTTCATCCAGGACGAATGTGGTGTCATTGCCGGA 984

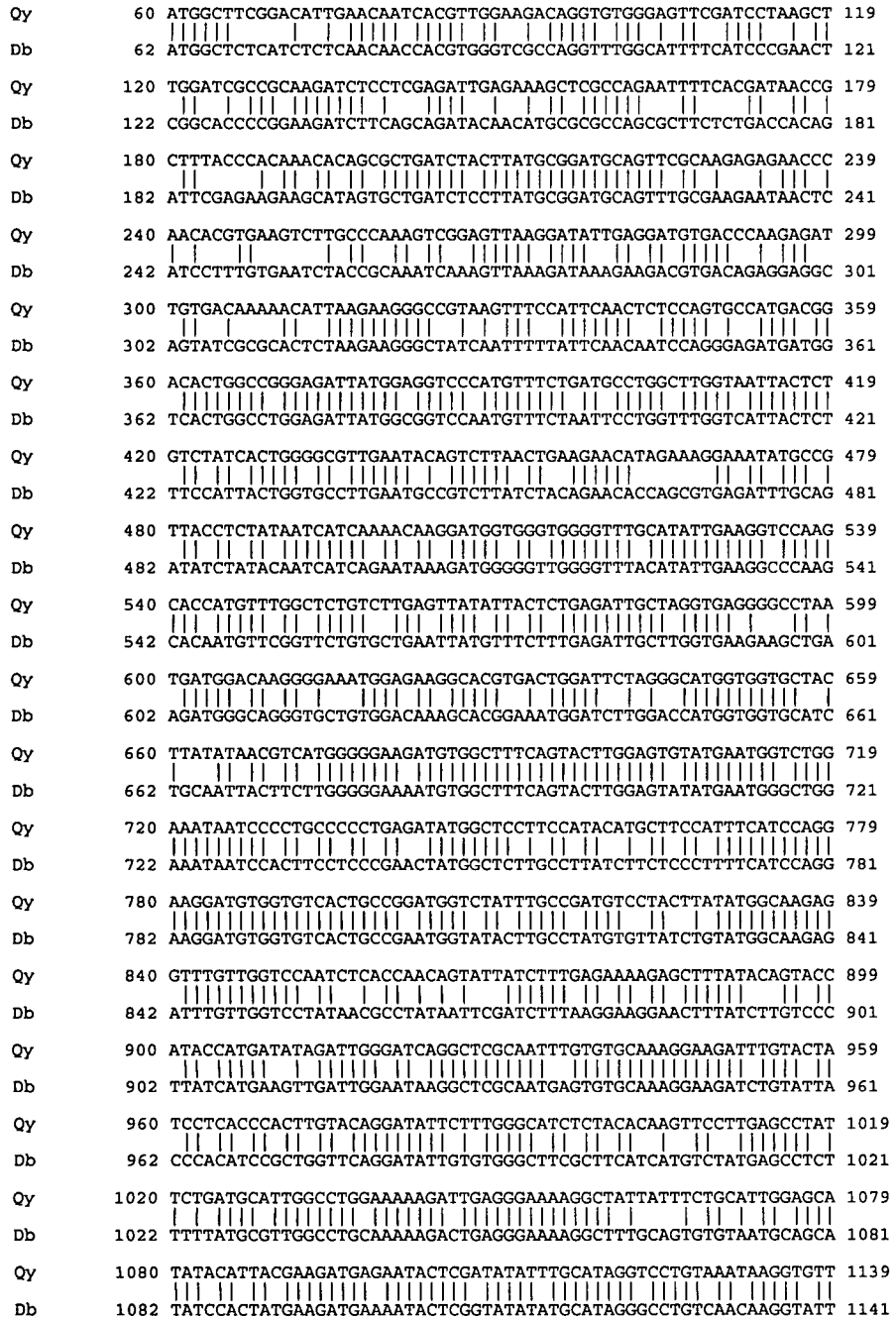
Qy        803 TGGTCTATTTGCCGATGTCCTACTTATATGGCAAGAGGTTTGTTGGTCCAATCTCACCAA 862
Db        985 TGGTTTACCTGCCCATGTCATACATCTATGGAAAGAGGTTTACTGGTCCAATTACACCAC 1044

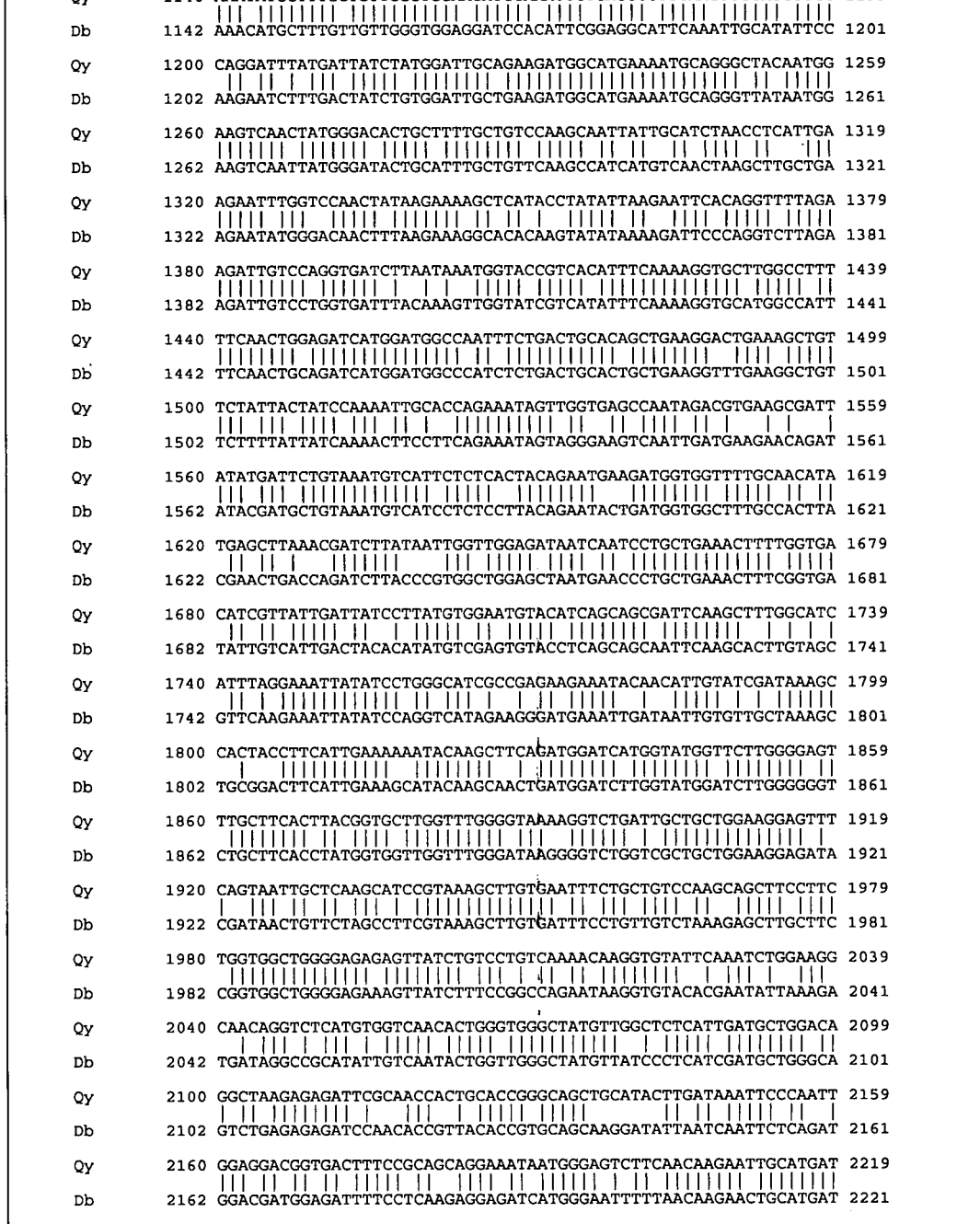
Qy        863 CAGTATTATCTTTGAGAAAAGAGCTTTATACAGTACCATACCATGATATAGATTGGGATC 922
Db       1045 TTATCTTGTCACTAAGGAAGGAACTATTTAATCTTCCATTTGATCAGTTAGATTGGAACA 1104

Qy        923 AGGCTCGCAATTTGTGTGCAAAGGAAGATTTGTACTATCCTCACCCACTTGTACAGGATA 982
Db       1105 AGGCTCGTAACGAGTGTGCAAAGGAAGATCTGTACTATCCACATCCATTTATACAAGACG 1164

Qy        983 TTCTTTGGGCATCTCTACACAAGTTCCTTGAGCCTATTCTGATGCATTGGCCTGGAAAAA 1042
Db       1165 TCCTTTGGGCATCTCTACATAAATTTGTCGAACCTATTCTGATGCATTGGCCTGGAAGCA 1224

Qy       1043 GATTGAGGGAAAAGGCTATTATTTCTGCATTGGAGCATATACATTACGAAGATGAGAATA 1102
Db       1225 AATTAAGAGAGAAGGCTGTTAATACTGCTATGCAGCATGTCCATTATGAGGATGAAAATA 1284
```





```
Qy       1080 TATACATTACGAAGATGAGAATACTCGATATATTTGCATAGGTCCTGTAAATAAGGTGTT 1139
Db       1202 TATCCACTATGAAGATGAAAATACTCGGTATATATGCATAGGGCCTGTCAACAAGGTGTT 1261

Qy       1140 AAATATGCTTTGCTGTTGGGTGGAAGATCCAAATTCTGAGGCCTTCAAGTTGCATCTTCC 1199
Db       1262 AAACATGCTTTGCTGTTGGGCGGAAGATCCACATTCAGAGGCATTCAAATTGCATATTCC 1321

Qy       1200 CAGGATTTATGATTATCTATGGATTGCAGAAGATGGCATGAAAATGCAGGGCTACAATGG 1259
Db       1322 TAGAATCTATGACTATCTGTGGATTGCTGAAGATGGCATGAAAATGCAGGGTTATAATGG 1381

Qy       1260 AAGTCAACTATGGGACACTGCTTTTGCTGTCCAAGCAATTATTGCATCTAACCTCATTGA 1319
Db       1382 AAGTCAATTATGGGATACTGCATTTGCTGTTCAAGCCATCATTTCAACTGAGCTCGCTGA 1441

Qy       1320 AGAATTTGGTCCAACTATAAGAAAAGCTCATACCTATATTAAGAATTCACAGGTTTTAGA 1379
Db       1442 AGAATATGAGACAACTTTAAGAAAGGCGCACAAGTATATAAAAGACTCCCAGGTCTTAGA 1501

Qy       1380 AGATTGTCCAGGTGATCTTAATAAATGGTACCGTCACATTTCAAAAGGTGCTTGGCCTTT 1439
Db       1502 AGATTGTCCTGGAGATTTACAAAGTTGGTACCGTCATATTTCAAAAGGTGCATGGCCATT 1561

Qy       1440 TTCAACTGGAGATCATGGATGGCCAATTTCTGACTGCACAGCTGAAGGACTGAAAGCTGT 1499
Db       1562 TTCAACTGCGGATCACGGATGGCCCATCTCTGACTGCACTGCTGAAGGTTTGAAGGCTGT 1621

Qy       1500 TCTATTACTATCCAAAATTGCACCAGAAATAGTTGGTGAGCCAATAGACGTGAAGCGATT 1559
Db       1622 TCTTTTATTGTCAAAACTTCCTTCAGAAATAGTAGGGAAGTCAATTGATGAACAACAGCT 1681

Qy       1560 ATATGATTCTGTAAATGTCATTCTCTCACTACAGAATGAAGATGGTGGTTTTGCAACATA 1619
Db       1682 ATACAACGCAGTAAATGTCATCCTGTCCTTACAGAATACTGATGGTGGCTTTGCAACTTA 1741

Qy       1620 TGAGCTTAAACGATCTTATAATTGGTTGGAGATAATCAATCCTGCTGAAACTTTTGGTGA 1679
Db       1742 TGAGTTGACCAGGTCATATCGGTGGCTGGAGCTAATGAACCCTGCTGAAACTTTTGGCGA 1801

Qy       1680 CATCGTTATTGATTATCCTTATGTGGAATGTACATCAGCAGCGATTCAAGCTTTGGCATC 1739
Db       1802 TATTGTCATTGACTACCCATATGTCGAGTGTTCCTCAGCAGCGATTCAAGCACTTGCAGC 1861

Qy       1740 ATTTAGGAAATTATATCCTGGGCATCGCCGAGAAGAAATACAACATTGTATCGATAAAGC 1799
Db       1862 ATTCAAGAAATTATATCCAGGTCATAGAAGGGATGAAATTGATAATTGTATTGCTGAAGC 1921

Qy       1800 CACTACCTTCATTGAAAAAATACAAGCTTCAGATGGATCATGGTATGGTTCTTGGGGAGT 1859
Db       1922 TGCGGACTTCATTGAAAGTATACAAGCAACTGATGGATCTTGGTATGGATCTTGGGGAGT 1981

Qy       1860 TTGCTTCACTTACGGTGCTTGGTTTGGGGTAAAAGGTCTGATTGCTGCTGGAAGGAGTTT 1919
Db       1982 CTGCTTCACCTATGGTGGTTGGTTCGGGATAAGGGGGCTGGTCGCTGCTGGAAGGAGATA 2041

Qy       1920 CAGTAATTGCTCAAGCATCCGTAAAGCTTGTGAATTTCTGCTGTCCAAGCAGCTTCCTTC 1979
Db       2042 CAATAACTGTTCTAGCCTTCGTAAAGCTTGTGATTTCCTGTTGTCCAAAGAGCTTGCTGC 2101

Qy       1980 TGGTGGCTGGGGAGAGAGTTATCTGTCCTGTCAAAACAAGGTGTATTCAAATCTGGAAGG 2039
Db       2102 CGGTGGCTGGGGAGAAAGTTACCTTTCCTGCCAGAATAAGGTGTACACAAATATTAAAGA 2161

Qy       2040 CAACAGGTCTCATGTGGTCAACACTGGGTGGGCTATGTTGGCTCTCATTGATGCTGGACA 2099
Db       2162 TGACAGGCCACATATTGTCAATACTGGTTGGGCTATGTTATCCCTCATCGATGCTGGGCA 2221

Qy       2100 GGCTAAGAGAGATTCGCAACCACTGCACCGGGCAGCTGCATACTTGATAAATTCCCAATT 2159
Db       2222 GTCTGAGAGAGATCCAACACCATTACACCGTGCAGCAAGGGTATTAATTAATTCTCAGAT 2281

Qy       2160 GGAGGACGGTGACTTTCCGCAGCAGGAAATAATGGGAGTCTTCAACAAGAATTGCATGAT 2219
Db       2282 GGAAGACGGAGATTTTCCTCAAGAGGAGATCATGGGAGTCTTTAACAAGAACTGCATGAT 2341

Qy       2220 CACATACGCCGCATACAGAAACATATTCCCCATTTGGGCGTTGGGAGAATACCAATCCCA 2279
Db       2342 TAGCTATTCAGCATACCGTAACATTTTCCCTATATGGGCCCTTGGGGAATACCGTAGCCG 2401

Qy       2280 AGTATTGCAATCTCGTTAATCG 2301
Db       2402 GGTACTGAAACCCCTTAAATAG 2423
```

RESULT 13
AB033334
LOCUS AB033334 2433 bp mRNA linear PLN 07-JAN-2000
DEFINITION Luffa cylindrica LcCAS1 mRNA for cycloartenol synthase, complete cds.
ACCESSION AB033334
VERSION AB033334.1 GI:6045132
KEYWORDS cycloartenol synthase.
SOURCE Luffa cylindrica (smooth loofah)
ORGANISM Luffa cylindrica
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Cucurbitales; Cucurbitaceae; Luffa.
REFERENCE 1 (sites)
AUTHORS Hayashi,H., Hiraoka,N., Ikeshiro,Y., Yazaki,K., Tanaka,S., Kushiro,T., Shibuya,M. and Ebizuka,Y.
TITLE Molecular cloning of a cDNA encoding cycloartenol synthase from Luffa cylindrica
JOURNAL Plant Physiol. 121, 1383 (1999)
REFERENCE 2 (bases 1 to 2433)
AUTHORS Hayashi,H.
TITLE Direct Submission
JOURNAL Submitted (06-OCT-1999) Hiroaki Hayashi, Gifu Pharmaceutical University, Department of Pharmacognosy; 5-6-1 Mitahora-higashi, Gifu, Gifu 502-8585, Japan (E-mail:hayashih@gifu-pu.ac.jp, Tel:81-58-237-3931(ex.325), Fax:81-58-237-5979)

FEATURES Location/Qualifiers
source 1. .2433
/organism="Luffa cylindrica"
/mol_type="mRNA"
/db_xref="taxon:3670"
/cell_type="suspension cultured cells"
/clone_lib="lambda ZAPII"
gene 1. .2433
/gene="LcCAS1"
CDS 3. .2300
/gene="LcCAS1"
/codon_start=1
/product="cycloartenol synthase"
/protein_id="BAA85266.1"
/db_xref="GI:6045133"
/translation="MWQLKIGADTVPADPSNAGGWLSSLNNHVGRQVWHFHPELGTPE
DLQQIQHARQRFSDHRFEKKHSADLLMRMQFAKNNSSFVNLPQIKVKDKEDVTEEAVS
RTLRRAINFYSTIQGDDGHWPGDYGGPMFLIPGLVITLSITGALNAVLSTEHQREICR
YLYNHQNKDGGWGLHIEGPSTMFGSVLNYVSLRLLGEEAEDGQGAVDKARKWILDHGG
ASAITSWGKMWLSVLGVYEWAGNNPLPPELWLLPYLLPFHPGRMWCHCRMVYLPMCYL
YGKRFVGPITPIIRSLRKELYLVPYHEVDWNKARNECAKEDLYYPHPLVQDIVWASLH
HVYEPLFMRWPAKRLREKALQCVMQHIHYEDENTRYICIGPVNKVLNMLCCWVEDPHS
EAFKLHIPRIFDYLWIAEDGMKMQGYNGSQLWDTAFAVQAIMSTKLAEEYGTTLRKAH
KYIKDSQVLEDCPGDLQSWYRHISKGAWPFSTADHGWPISDCTAEGLKAVLLLSKLPS
EIVGKSIDEEQIYDAVNVILSLQNTDGGFATYELTRSYPWLELMNPAETFGDIVIDYT
YVECTSAAIQALVAFKKLYPGHRRDEIDNCVAKAADFIESIQATDGSWYGSWGVCFTY
GGWFGIRGLVAAGRRYDNCSSLRKACDFLLSKELASGGWGESYLSGQNKVYTNIKDDR
PHIVNTGWAMLSLIDAGQSERDPTPLHRAARILINSQMDDGDFPQEEIMGIFNKNCMI
SYAAYRNIFPIWALGEYRCRVLQAP"

ORIGIN

Query Match 52.8%; Score 1351; DB 8; Length 2433;
Best Local Similarity 75.2%; Pred. No. 0;
Matches 1684; Conservative 0; Mismatches 555; Indels 0; Gaps 0;

```
Db       2075 GTATACCAATCTCAAAGATAATAGGCCACATATTGTCAATACTGGATGGGCTATGTTGGC 2134

Qy       2082 TCTCATTGATGCTGGACAGGCTAAGAGAGATTCGCAACCACTGCACCGGGCAGCTGCATA 2141

Db       2135 CCTCATTGATGCCGGACAGGCTGAGAGAGATCCAACCCCATTGCACCGAGCGGCAAGAAT 2194

Qy       2142 CTTGATAAATTCCCAATTGGAGGACGGTGACTTTCCGCAGCAGGAAATAATGGGAGTCTT 2201

Db       2195 ACTAATAAATTCTCAAATGGAGAATGGAGATTTTCCTCAAGAGGAGATTATGGGAGTTTT 2254

Qy       2202 CAACAAGAATTGCATGATCACATACGCCGCATACAGAAACATATTCCCCATTTGGGCGTT 2261

Db       2255 CAACAAGAATTGTATGATCAGCTATTCTGCATACCGAAACATTTTTCCTATCTGGGCCCT 2314

Qy       2262 GGGAGAATACCAATCCCAAGTATTGCAATCTCGTTAAT 2299

Db       2315 TGGAGAATATCGTTGTCGGGTATTGAAGGCCCTCTGAT 2352
```

RESULT 12
AB116237
LOCUS AB116237 2556 bp mRNA linear PLN 03-AUG-2004
DEFINITION Cucurbita pepo CPX mRNA for cycloartenol synthase, complete cds.
ACCESSION AB116237
VERSION AB116237.1 GI:50896400
KEYWORDS .
SOURCE Cucurbita pepo
ORGANISM Cucurbita pepo
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Cucurbitales; Cucurbitaceae; Cucurbita.
REFERENCE 1
AUTHORS Shibuya,M., Adachi,S. and Ebizuka,Y.
TITLE Cucurbitadienol Synthase, the First Committed Enzyme for Cucurbitacin Biosynthesis, is a Distinct Enzyme from Cycloartenol Synthase for Phytosterol Biosynthesis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2556)
AUTHORS Shibuya,M.
TITLE Direct Submission
JOURNAL Submitted (01-AUG-2003) Masaaki Shibuya, The University of Tokyo, Graduate School of Pharmaceutical Sciences; Hongo 7-3-1, Bunkyo-ku, Tokyo 1130033, Japan (E-mail:shibuyam@mol.f.u-tokyo.ac.jp, URL:http://www.f.u-tokyo.ac.jp/~tennen/head.htm, Tel:81-3-5841-4741, Fax:81-3-5841-4744)
FEATURES Location/Qualifiers
source 1. .2556
/organism="Cucurbita pepo"
/mol_type="mRNA"
/db_xref="taxon:3663"
gene 1. .2556
/gene="CPX"
CDS 123. .2423
/gene="CPX"
/codon_start=1
/product="cycloartenol synthase"
/protein_id="BAD34644.1"
/db_xref="GI:50896401"
/translation="MWQLKIGADTVPSDPSNAGGWLSTLNNHVGRQVWHFHPELGSPE
DLQQIQQARQHFSDHRFEKKHSADLLMRMQFAKENSSFVNLPQVKVKDKEDVTEEAVT
RTLRRAINFYSTIQADDGHWPGDYGGPMFLIPGLVITLSITGALNAVLSTEHQREICR
YLYNHQNKDGGWGLHIEGPSTMFGSVLNYVTLRLLGEEAEDGQGAVDKARKWILDHGG
AAAITSWGKMWLSVLGVYEWAGNNPLPPELWLLPYLLPCHPGRMWCHCRMVYLPMCYL
YGKRFVGPITPIIRSLRKELYLVPYHEVDWNKARNQCAKEDLYYPHPLVQDILWATLH
HVYEPLFMHWPAKRLREKALQSVMQHIHYEDENTRYICIGPVNKVLNMLCCWAEDPHS
EAFKLHIPRIYDYLWIAEDGMKMQGYNGSQLWDTAFAVQAIISTELAEEYETTLRKAH
KYIKDSQVLEDCPGDLQSWYRHISKGAWPFSTADHGWPISDCTAEGLKAVLLLSKLPS
EIVGKSIDEQQLYNAVNVILSLQNTDGGFATYELTRSYRWLELMNPAETFGDIVIDYP
YVECSSAAIQALAAFKKLYPGHRRDEIDNCIAEAADFIESIQATDGSWYGSWGVCFTY
GGWFGIRGLVAAGRRYNNCSSLRKACDFLLSKELAAGGWGESYLSCQNKVYTNIKDDR
PHIVNTGWAMLSLIDAGQSERDPTPLHRAARVLINSQMEDGDFPQEEIMGVFNKNCMI
SYSAYRNIFPIWALGEYRSRVLKPLK"

```
ORIGIN

  Query Match           52.8%;  Score 1352.4;  DB 8;  Length 2556;
  Best Local Similarity 75.2%;  Pred. No. 0;
  Matches 1686;  Conservative    0;  Mismatches  556;  Indels    0;  Gaps    0;

Qy         60 ATGGCTTCGGACATTGAACAATCACGTTGGAAGACAGGTGTGGGAGTTCGATCCTAAGCT 119

Db        182 ATGGCTTTCAACTCTCAACAACCACGTCGGCCGCCAAGTTTGGCATTTTCATCCCGAACT 241

Qy        120 TGGATCGCCGCAAGATCTCCTCGAGATTGAGAAAGCTCGCCAGAATTTTCACGATAACCG 179

Db        242 CGGCTCCCCGGAGGATCTTCAGCAAATACAACAGGCTCGTCAGCACTTCTCCGATCACAG 301

Qy        180 CTTTACCCACAAACACAGCGCTGATCTACTTATGCGGATGCAGTTCGCAAGAGAGAACCC 239

Db        302 ATTCGAGAAGAAGCATAGTGCTGATCTCCTCATGCGAATGCAGTTTGCGAAGGAGAATTC 361

Qy        240 AACACGTGAAGTCTTGCCCAAAGTCGGAGTTAAGGATATTGAGGATGTGACCCAAGAGAT 299

Db        362 GTCCTTTGTGAATCTACCGCAAGTCAAAGTTAAAGATAAAGAAGACGTGACAGAGGAAGC 421

Qy        300 TGTGACAAAAACATTAAGAAGGGCCGTAAGTTTCCATTCAACTCTCCAGTGCCATGACGG 359

Db        422 AGTAACGCGCACTCTAAGAAGGGCTATCAATTTTTATTCAACAATCCAGGCAGATGACGG 481

Qy        360 ACACTGGCCGGGAGATTATGGAGGTCCCATGTTTCTGATGCCTGGCTTGGTAATTACTCT 419

Db        482 TCATTGGCCTGGTGATTATGGCGGCCCAATGTTTCTAATTCCTGGTTTGGTCATTACTCT 541

Qy        420 GTCTATCACTGGGGCGTTGAATACAGTCTTAACTGAAGAACATAGAAAGGAAATATGCCG 479

Db        542 TTCCATWACTGGCGCCTTGAATGCGGTCTTATCTACAGAACACCAGCGTGAGATTTGCAG 601

Qy        480 TTACCTCTATAATCATCAAAACAAGGATGGTGGGTGGGGTTTGCATATTGAAGGTCCAAG 539

Db        602 ATATCTGTACAATCATCAGAATAAAGATGGGGGTTGGGGTTTACATATTGAGGGCCCAAG 661

Qy        540 CACCATGTTTGGCTCTGTCTTGAGTTATATTACTCTGAGATTGCTAGGTGAGGGGCCTAA 599

Db        662 CACAATGTTCGGATCTGTGTTGAATTATGTTACTTTGAGATTGCTTGGTGAAGAAGCTGA 721

Qy        600 TGATGGACAAGGGGAAATGGAGAAGGCACGTGACTGGATTCTAGGGCATGGTGGTGCTAC 659

Db        722 AGATGGGCAGGGTGCTGTGGACAAAGCACGGAAATGGATCTTGGACCATGGTGGTGCAGC 781

Qy        660 TTATATAACGTCATGGGGGAAGATGTGGCTTTCAGTACTTGGAGTGTATGAATGGTCTGG 719

Db        782 TGCGATTACTTCTTGGGGGAAAATGTGGCTTTCAGTACTTGGAGTATATGAATGGGCTGG 841

Qy        720 AAATAATCCCCTGCCCCCTGAGATATGGCTCCTTCCATACATGCTTCCATTTCATCCAGG 779

Db        842 AAATAATCCACTTCCTCCAGAGCTATGGCTCTTGCCTTATCTTCTCCCTTGTCATCCAGG 901

Qy        780 AAGGATGTGGTGTCACTGCCGGATGGTCTATTTGCCGATGTCCTACTTATATGGCAAGAG 839

Db        902 AAGGATGTGGTGTCACTGTCGAATGGTATACTTGCCTATGTGTTATCTGTATGGCAAGAG 961

Qy        840 GTTTGTTGGTCCAATCTCACCAACAGTATTATCTTTGAGAAAAGAGCTTTATACAGTACC 899

Db        962 ATTTGTTGGTCCTATAACGCCTATAATTAGATCTTTAAGGAAGGAACTTTATCTTGTTCC 1021

Qy        900 ATACCATGATATAGATTGGGATCAGGCTCGCAATTTGTGTGCAAAGGAAGATTTGTACTA 959

Db       1022 TTATCATGAAGTTGATTGGAATAAGGCTCGAAATCAATGTGCAAAGGAAGATCTATATTA 1081

Qy        960 TCCTCACCCACTTGTACAGGATATTCTTTGGGCATCTCTACACAAGTTCCTTGAGCCTAT 1019

Db       1082 CCCACATCCGCTGGTTCAGGATATTCTGTGGGCTACGCTGCACCATGTCTATGAGCCTCT 1141

Qy       1020 TCTGATGCATTGGCCTGGAAAAAGATTGAGGGAAAAGGCTATTATTTCTGCATTGGAGCA 1079

Db       1142 TTTTATGCATTGGCCTGCAAAAAGATTGAGGGAGAAGGCTTTGCAGAGTGTAATGCAGCA 1201
```

NSEAFKLHLPRILDYLWIAEDGMKMQGYNGSQLWDTTFAVQAIISTNIAEEYGQTLRK
AHEYIKDSQVLEDCPGDLNFWYRHISKGAWPFSTADHGWPISDCTAEGLKAVILLSQF
PSETVGKSVDVKRLYDAVHVILSLQNTDGGFATYELTRSYHWLELINPAETFGDIVID
YPYVECTSAAIQALTLFKKLHPGHRREEIENCIAKAAEFIENIQASDGSWYGSWGVCF
TYAGWFGIKGLVAAGRTYKNCSSIHKACDYLLSKELASGGWGESYLSCQDKVYTNLKD
NRPHIVNTGWAMLALIDAGQAERDPTPLHRAARILINSQMENGDFPQEEIMGVFNKNC
MISYSAYRNIFPIWALGEYRCRVLKAL"

ORIGIN

Query Match 54.5%; Score 1394; DB 8; Length 2460;
Best Local Similarity 76.4%; Pred. No. 0;
Matches 1725; Conservative 0; Mismatches 530; Indels 3; Gaps 1;

```
Qy      42 CGCCGAAGGAGGGAATCCATGGCTTCGGACATTGAACAATCACGTTGGAAGACAGGTGTG 101
Db      98 CGGCGGCGGCAGTGAAACATGGCTGAGGAGTTTGAACAATCACTTGGGACGCCAAATATG 157

Qy     102 GGAGTTCGATCCTAAGCTTGGATCGCCGCAAGATCTCCTCGAGATTGAGAAAGCTCGCCA 161
Db     158 GGAGTTCCATCCAGAGCTCGGCACCCAAGAAGAGCTTCAGCAGATCGACGATGCTCGCCG 217

Qy     162 GAATTTTCACGATAACCGCTTTACCCACAAACACAGCGCTGATCTACTTATGCGGATGCA 221
Db     218 GAGGTTCTGGGAACGCCGCTTCGAAAGGCGGCACAGCTCCGACCTCCTCATGCGGATTCA 277

Qy     222 GTTCGCAAGAGAGAACCCAACACGTGAAGTCTTGCCCAAAGTCGGAGTTAAGGATATTGA 281
Db     278 GTTTGCAAAGGAGAATCCATCTTCTGCAAATATACCACAAGTCAAAATTAAAGATACAGA 337

Qy     282 GGATGTGACCCAAGAGATTGTGACAAAAACATTAAGAAGGGCCGTAAGTTTCCATTCAAC 341
Db     338 GGAGGTCAGAGAGGAGGCTGTAGGAATGACCTTACGAAGGGCCATCAATTTCTACTCAAC 397

Qy     342 TCTCCAGTGCCATGACGGACACTGGCCGGGAGATTATGGAGGTCCCATGTTTCTGATGCC 401
Db     398 AATACAGGCAGATGATGGTCATTGGCCTGGGGATTATGGGGGTCCCATGTTTCTAATTCC 457

Qy     402 TGGCTTGGTAATTACTCTGTCTATCACTGGGGCGTTGAATACAGTCTTAACTGAAGAACA 461
Db     458 TGGTCTGGTGATTACTCTTTCAATTACTGGAACTCTAAATGCTTTCTTATCAAAGGAACA 517

Qy     462 TAGAAAGGAAATATGCCGTTACCTCTATAATCATCAAAACAAGGATGGTGGGTGGGGTTT 521
Db     518 TCAATGTGAGATCTGCAGATATCTCTACAATCATCAGAATGAAGATGGGGGTTGGGGTCT 577

Qy     522 GCATATTGAAGGTCCAAGCACCATGTTTGGCTCTGTCTTGAGTTATATTACTCTGAGATT 581
Db     578 ACATATTGAGGGCCCAAGCACCATGTTTGGTACTGCACTAAATTATATTACTTTGAGATT 637

Qy     582 GCTAGGTGAGGGGCCTAATGATGGACAAGGGGAAATGGAGAAGGCACGTGACTGGATTCT 641
Db     638 GCTTGGTGA---ACCTGAGGATGGAATGGGAGCTGTGGAAAAAGCACGGAAGTGGATCTT 694

Qy     642 AGGGCATGGTGGTGCTACTTATATAACGTCATGGGGGAAGATGTGGCTTTCAGTACTTGG 701
Db     695 GGATCATGGTGGTGCAACTGCAATCACTTCATGGGGAAAAATGTGGCTTTCAGTACTTGG 754

Qy     702 AGTGTATGAATGGTCTGGAAATAATCCCCTGCCCCCTGAGATATGGCTCCTTCCATACAT 761
Db     755 AGTGTACGAATGGTCTGGGAATAATCCACTGCCACCAGAAGTATGGCTTTGCCCTTATCT 814

Qy     762 GCTTCCATTTCATCCAGGAAGGATGTGGTGTCACTGCCGGATGGTCTATTTGCCGATGTC 821
Db     815 TCTCCCATGTCATCCAGGAAGGATGTGGTGTCATTGCCGAATGGTGTATTTGCCCATGTC 874

Qy     822 CTACTTATATGGCAAGAGGTTTGTTGGTCCAATCTCACCAACAGTATTATCTTTGAGAAA 881
Db     875 ATATTTATATGGTAAGAGGTTTGTTGGTCCTATCACATCCACAATCCAATCTTTGAGAAA 934

Qy     882 AGAGCTTTATACAGTACCATACCATGATATAGATTGGGATCAGGCTCGCAATTTGTGTGC 941
Db     935 GGAGCTTTATACAGTCCCTTATCATGAGATCGATTGGAATAAGGCTCGCAATGATTGTGC 994

Qy     942 AAAGGAAGATTTGTACTATCCTCACCCACTTGTACAGGATATTCTTTGGGCATCTCTACA 1001
Db     995 AAAGGAAGATCTATACTACCCCCATCCACTGGTTCAGGATATCCTCTGGGCATCTCTTTA 1054

Qy    1002 CAAGTTCCTTGAGCCTATTCTGATGCATTGGCCTGGAAAAAGATTGAGGGAAAAGGCTAT 1061
Db    1055 CTATGCATATGAACCCATTTTTATGTATTGGCCTGCAAAAAGATTGAGGGAGAAGGCTCT 1114

Qy    1062 TATTTCTGCATTGGAGCATATACATTACGAAGATGAGAATACTCGATATATTTGCATAGG 1121
Db    1115 GGACACTGTGATGCAGCATATCCACTATGAAGATGAAAATACACGGTATATATGCATAGG 1174

Qy    1122 TCCTGTAAATAAGGTGTTAAATATGCTTTGCTGTTGGGTGGAAGATCCAAATTCTGAGGC 1181
Db    1175 GCCTGTCAATAAGGTGTTAAATATGCTTTGTTGTTGGGCTGAGGATCCAAACTCAGAGGC 1234

Qy    1182 CTTCAAGTTGCATCTTCCCAGGATTTATGATTATCTATGGATTGCAGAAGATGGCATGAA 1241
Db    1235 ATTCAAATTACATCTTCCAAGAATCTTAGATTATTTGTGGATTGCCGAAGATGGCATGAA 1294

Qy    1242 AATGCAGGGCTACAATGGAAGTCAACTATGGGACACTGCTTTTGCTGTCCAAGCAATTAT 1301
Db    1295 AATGCAGGGATATAATGGAAGTCAATTATGGGATACTACTTTTGCTGTTCAAGCGATTAT 1354

Qy    1302 TGCATCTAACCTCATTGAAGAATTTGGTCCAACTATAAGAAAAGCTCATACCTATATTAA 1361
Db    1355 ATCAACTAATATTGCTGAAGAATATGGACAAACTTTAAGAAAAGCACATGAGTACATAAA 1414

Qy    1362 GAATTCACAGGTTTTAGAAGATTGTCCAGGTGATCTTAATAAATGGTACCGTCACATTTC 1421
Db    1415 GGATTCCCAGGTCTTAGAAGATTGTCCAGGAGATCTAAATTTTTGGTATCGTCATATTTC 1474

Qy    1422 AAAAGGTGCTTGGCCTTTTTCAACTGGAGATCATGGATGGCCAATTTCTGACTGCACAGC 1481
Db    1475 AAAGGGTGCTTGGCCATTCTCAACTGCAGATCATGGATGGCCTATTTCTGACTGTACAGC 1534

Qy    1482 TGAAGGACTGAAAGCTGTTCTATTACTATCCAAAATTGCACCAGAAATAGTTGGTGAGCC 1541
Db    1535 GGAAGGATTGAAGGCTGTTATATTGCTATCACAATTTCCATCAGAAACAGTTGGGAAATC 1594

Qy    1542 AATAGACGTGAAGCGATTATATGATTCTGTAAATGTCATTCTCTCACTACAGAATGAAGA 1601
Db    1595 GGTAGATGTGAAGCGGTTATATGATGCTGTACATGTTATCCTTTCCTTACAGAATACTGA 1654

Qy    1602 TGGTGGTTTTGCAACATATGAGCTTAAACGATCTTATAATTGGTTGGAGATAATCAATCC 1661
Db    1655 TGGTGGCTTTGCAACTTATGAATTGACCAGATCCTACCATTGGTTGGAGCTAATCAATCC 1714

Qy    1662 TGCTGAAACTTTTGGTGACATCGTTATTGATTATCCTTATGTGGAATGTACATCAGCAGC 1721
Db    1715 TGCTGAAACTTTTGGTGATATTGTCATTGATTACCCATATGTTGAATGTACATCAGCAGC 1774

Qy    1722 GATTCAAGCTTTGGCATCATTTAGGAAATTATATCCTGGGCATCGCCGAGAAGAAATACA 1781
Db    1775 AATTCAAGCACTGACATTATTCAAGAAATTACATCCTGGGCATAGGAGGGAAGAAATAGA 1834

Qy    1782 ACATTGTATCGATAAAGCCACTACCTTCATTGAAAAAATACAAGCTTCAGATGGATCATG 1841
Db    1835 AAATTGTATTGCTAAAGCTGCAGAGTTTATTGAAAATATACAAGCATCAGATGGCTCTTG 1894

Qy    1842 GTATGGTTCTTGGGGAGTTTGCTTCACTTACGGTGCTTGGTTTGGGGTAAAAGGTCTGAT 1901
Db    1895 GTATGGATCCTGGGGAGTCTGTTTCACCTATGCTGGCTGGTTTGGGATAAAGGGGTTGGT 1954

Qy    1902 TGCTGCTGGAAGGAGTTTCAGTAATTGCTCAAGCATCCGTAAAGCTTGTGAATTTCTGCT 1961
Db    1955 TGCTGCTGGAAGGACATACAAGAATTGTTCTAGCATCCACAAGGCATGCGATTATCTGTT 2014

Qy    1962 GTCCAAGCAGCTTCCTTCTGGTGGCTGGGGAGAGAGTTATCTGTCCTGTCAAAACAAGGT 2021
Db    2015 GTCCAAAGAGCTCGCTTCAGGTGGTTGGGGAGAAAGTTACCTATCATGTCAGGACAAGGT 2074

Qy    2022 GTATTCAAATCTGGAAGGCAACAGGTCTCATGTGGTCAACACTGGGTGGGCTATGTTGGC 2081
```

```
Qy        865 GTATTATCTTTGAGAAAAGAGCTTTATACAGTACCATACCATGATATAGATTGGGATCAG 924
              || |||||  ||||||| |||||||  |||||||||||| |||||  |  | ||| || |
Db        841 GTTTTATCACTGAGAAAGGAGCTTTTCACAGTACCATATCATGAAGTCAACTGGAATGAA 900

Qy        925 GCTCGCAATTTGTGTGCAAAGGAAGATTTGTACTATCCTCACCCACTTGTACAGGATATT 984
              || |||||  | || ||||||| ||||| ||||| || || ||||||||| || |||||||
Db        901 GCACGCAACCTTTGCGCAAAGGAGGATTTATACTACCCACATCCACTTGTGCAAGATATT 960

Qy        985 CTTTGGGCATCTCTACACAAGTTCCTTGAGCCTATTCTGATGCATTGGCCTGGAAAAAGA 1044
              |||||||||||  || || |||  |  ||||||||  |||||||||   ||||||||   ||
Db        961 CTTTGGGCATCACTTCATAAGATTGTTGAGCCTGTTCTGATGCGATGGCCTGGTGCAAAT 1020

Qy       1045 TTGAGGGAAAAGGCTATTATTTCTGCATTGGAGCATATACATTACGAAGATGAGAATACT 1104
              |||||  ||||||||||| |    | ||| | || |||||  ||||| ||||||||||||||||
Db       1021 TTGAGAGAAAAGGCTATAAGAACCGCAATAGAACATATTCATTATGAAGATGAGAATACT 1080

Qy       1105 CGATATATTTGCATAGGTCCTGTAAATAAGGTGTTAAATATGCTTTGCTGTTGGGTGGAA 1164
              | || || ||||||||||| || || |||||  |||||||||||||||||||||||||| |||
Db       1081 AGGTACATCTGCATAGGTCCCGTGAACAAGGTATTAAATATGCTTTGCTGTTGGGTAGAA 1140

Qy       1165 GATCCAAATTCTGAGGCCTTCAAGTTGCATCTTCCCAGGATTTATGATTATCTATGGATT 1224
              || |||||  || |||||  |||||||||||  || || ||  ||  |||| | ||| ||| |
Db       1141 GACCCAAACTCAGAGGCTTTCAAGTTGCACCTACCAAGAATCCATGACTTTCTCTGGTTA 1200

Qy       1225 GCAGAAGATGGCATGAAAATGCAGGGCTACAATGGAAGTCAACTATGGGACACTGCTTTT 1284
              || ||||||||  |||||  |||||||| || || |||||  || ||||||||  || | ||||
Db       1201 GCTGAAGATGGAATGAAGATGCAGGGTTATAACGGAAGCCAGCTATGGGATACAGGTTTT 1260

Qy       1285 GCTGTCCAAGCAATTATTGCATCTAACCTCATTGAAGAATTTGGTCCAACTATAAGAAAA 1344
              ||| | |||||  ||| | ||| ||||||||  | |||||||  ||| ||    | |   ||||
Db       1261 GCTATTCAAGCGATTTTGGCAACTAACCTCGTCGAAGAATATGGGCCCGTTTTGGAAAAA 1320

Qy       1345 GCTCATACCTATATTAAGAATTCACAGGTTTTAGAAGATTGTCCAGGTGATCTTAATAAA 1404
              || ||| | | | | |||||||| ||||| |||||||| || || || ||||| ||| |
Db       1321 GCACATTCATTTGTCAAGAATTCCCAGGTGTTAGAAGACTGCCCTGGAGATCTGAATTAC 1380

Qy       1405 TGGTACCGTCACATTTCAAAAGGTGCTTGGCCTTTTTCAACTGGAGATCATGGATGGCCA 1464
              |||||  || |||||||| |||||  |||||||||||  |||||||  |||||| || |||||
Db       1381 TGGTATCGCCACATTTCTAAAGGGGCTTGGCCTTTCTCAACTGCAGATCACGGTTGGCCC 1440

Qy       1465 ATTTCTGACTGCACAGCTGAAGGACTGAAAGCTGTTCTATTACTATCCAAAATTGCACCA 1524
              || |||||||||| || |||||||||||||||| ||| || |||||||||  || |
Db       1441 ATCTCTGACTGCACCGCAGAAGGACTGAAAGCTGCTCTTTTGCTATCCAAAGTTCCCAAG 1500

Qy       1525 GAAATAGTTGGTGAGCCAATAGACGTGAAGCGATTATATGATTCTGTAAATGTCATTCTC 1584
              |  || |||||||| |||||||| |  || || ||||||||  |||| |||||  ||   |
Db       1501 GCGATTGTTGGTGAACCAATAGATGCAAAACGGTTATATGAAGCTGTTAATGTTATCATT 1560

Qy       1585 TCACTACAGAATGAAGATGGTGGTTTTGCAACATATGAGCTTAAACGATCTTATAATTGG 1644
              ||  |||||||||  |||||| ||  | ||||||||||||||  |    | || ||   ||||
Db       1561 TCTTTACAGAATGCAGATGGAGGCCTCGCAACATATGAGCTCACCAGGTCATACCCTTGG 1620

Qy       1645 TTGGAGATAATCAATCCTGCTGAAACTTTTGGTGACATCGTTATTGATTATCCTTATGTG 1704
              || ||| ||||||| || || ||||| ||||| || || ||||||||||||||||||| |||
Db       1621 TTAGAGCTAATCAACCCAGCAGAAACCTTTGGCGATATTGTTATTGATTATCCTTACGTG 1680

Qy       1705 GAATGTACATCAGCAGCGATTCAAGCTTTGGCATCATTTAGGAAATTATATCCTGGGCAT 1764
              |||||||||||||| || || |||||||||  |||||||  | ||  | ||||||||  |||
Db       1681 GAATGTACATCAGCTGCTATCCAAGCTTTGATATCATTTCGAAAGCTGTATCCTGGTCAT 1740

Qy       1765 CGCCGAGAAGAAATACAACATTGTATCGATAAAGCCACTACCTTCATTGAAAAAATACAA 1824
              ||      | ||| || |  | || || || || ||    ||  |||||||||    || |||
Db       1741 CGAAAGAAGGAAGTAGATGAGTGCATTGAGAAGGCGGTTAAGTTCATTGAATCCATTCAA 1800

Qy       1825 GCTTCAGATGGATCATGGTATGGTTCTTGGGGAGTTTGCTTCACTTACGGTGCTTGGTTT 1884
              ||  ||||||| |||||||||| || ||||  |||||||||| || ||| | ||||||
Db       1801 GCAGCAGATGGCTCATGGTATGGATCATGGGCTGTTTGCTTCACGTATGGTACGTGGTTT 1860

Qy       1885 GGGGTAAAAGGTCTGATTGCTGCTGGAAGGAGTTTCAGTAATTGCTCAAGCATCCGTAAA 1944
              || || ||||| ||| | |||| |||||  |  || |  ||  |    ||    |  ||||
Db       1861 GGAGTGAAAGGGCTGGTAGCTGTTGGAAAAACATTGAAAAACTCTCCACATGTTGCTAAA 1920
```

```
Qy       1945 GCTTGTGAATTTCTGCTGTCCAAGCAGCTTCCTTCTGGTGGCTGGGGAGAGAGTTATCTG 2004
              ||||||||||||||  |||| || || |  ||||| || ||||||||||| || |||||
Db       1921 GCTTGTGAATTTCTATTGTCGAAACAACAACCTTCGGGCGGCTGGGGAGAAAGCTATCTT 1980

Qy       2005 TCCTGTCAAAACAAGGTGTATTCAAATCTGGAAGGCAACAGGTCTCATGTGGTCAACACT 2064
              || |||||| ||||||| |||||||| || || |||||||| ||||| || || || ||
Db       1981 TCATGTCAAGACAAGGTCTATTCAAACCTTGATGGCAACAGATCTCACGTCGTGAATACA 2040

Qy       2065 GGGTGGGCTATGTTGGCTCTCATTGATGCTGGACAGGCTAAGAGAGATTCGCAACCACTG 2124
              |  |||||||| | || ||||||| |||||| || ||| ||  |||   | |||||||
Db       2041 GCATGGGCTATGCTCGCACTCATTGGTGCTGGGCAAGCTGAGGTAGACCGGAAACCACTA 2100

Qy       2125 CACCGGGCAGCTGCATACTTGATAAATTCCCAATTGGAGGACGGTGACTTTCCGCAGCAG 2184
              |||||||| ||    ||||||||| ||| | ||| ||||| | ||||| ||||| || |||
Db       2101 CACCGGGCTGCAAGATACTTGATTAATGCTCAAATGGAGAATGGTGATTTTCCACAACAG 2160

Qy       2185 GAAATAATGGGAGTCTTCAACAAGAATTGCATGATCACATACGCCGCATACAGAAACATA 2244
              |||||||||||||||||||| | ||| |||||||| ||||| ||||| ||  |||||||
Db       2161 GAAATAATGGGAGTCTTCAATAGGAACTGCATGATAACATATGCCGCGTATCGAAACATT 2220

Qy       2245 TTCCCCATTTGGGCGTTGGGAGAATACCAATCCCAAGTATT 2285
              || || || ||||| ||||| || |||| |  || |||||
Db       2221 TTTCCGATATGGGCTTTGGGGGAGTACCGTTGTCAGGTATT 2261
```

```
RESULT 11
AB055509
LOCUS       AB055509               2460 bp    mRNA    linear   PLN 05-JAN-2002
DEFINITION  Betula platyphylla CASBPX1 mRNA for cycloartenol synthase, complete
            cds.
ACCESSION   AB055509
VERSION     AB055509.1  GI:18147589
KEYWORDS    .
SOURCE      Betula platyphylla
  ORGANISM  Betula platyphylla
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fagales; Betulaceae; Betula.
REFERENCE   1
  AUTHORS   Zhang,H., Shibuya,M., Ebizuka,Y. and Yokota,S.
  TITLE     Oxidosqualene Cyclases from Cell Suspension Culture of Betula
            platyphylla var. japonica:Molecular Evolution of Oxidosqualene
            Cyclases in Higher Plants
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2460)
  AUTHORS   Shibuya,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-FEB-2001) Masaaki Shibuya, University of Tokyo,
            Faculty of Pharmaceutical Sciences; 7-3-1 Hongo, Bunkyo-ku, Tokyo
            113-0033, Japan (E-mail:shibuyam@mol.f.u-tokyo.ac.jp,
            URL:http://www.f.u-tokyo.ac.jp/~tennen/head.htm,
            Tel:81-3-5841-4741, Fax:81-3-5841-4744)
FEATURES             Location/Qualifiers
     source          1. .2460
                     /organism="Betula platyphylla"
                     /mol_type="mRNA"
                     /db_xref="taxon:78630"
     gene            1. .2460
                     /gene="CASBPX1"
     CDS             48. .2351
                     /gene="CASBPX1"
                     /EC_number="5.4.99.8"
                     /codon_start=1
                     /product="cycloartenol synthase"
                     /protein_id="BAB83085.1"
                     /db_xref="GI:18147590"
                     /translation="MWKLKIGAETARGDGGGGGGSETWLRSLNNHLGRQIWEFHPELG
                     TQEELQQIDDARRRFWERRFERRHSSDLLMRIQFAKENPSSANIPQVKIKDTEEVREE
                     AVGMTLRRAINFYSTIQADDGHWPGDYGGPMFLIPGLVITLSITGTLNAFLSKEHQCE
                     ICRYLYNHQNEDGGWGLHIEGPSTMFGTALNYITLRLLGEPEDGMGAVEKARKWILDH
                     GGATAITSWGKMWLSVLGVYEWSGNNPLPPEVWLCPYLLPCHPGRMWCHCRMVYLPMS
                     YLYGKRFVGPITSTIQSLRKELYTVPYHEIDWNKARNDCAKEDLYYPHPLVQDILWAS
                     LYYAYEPIFMYWPAKRLREKALDTVMQHIHYEDENTRYICIGPVNKVLNMLCCWAEDP
```

```
Db       1981 TCATGTCAAGACAAGGTCTATTCAAACCTTGATGGCAACAGATCTCACGTCGTGAATACA 2040

Qy       2065 GGGTGGGCTATGTTGGCTCTCATTGATGCTGGACAGGCTAAGAGAGATTCGCAACCACTG 2124

Db       2041 GCATGGGCTATGCTCGCACTCATTGGTGCTGGGCAAGCTGAGGTAGACCGGAAACCACTA 2100

Qy       2125 CACCGGGCAGCTGCATACTTGATAAATTCCCAATTGGAGGACGGTGACTTTCCGCAGCAG 2184

Db       2101 CACCGGGCTGCAAGATACTTGATTAATGCTCAAATGGAGAATGGTGATTTTCCACAACAG 2160

Qy       2185 GAAATAATGGGAGTCTTCAACAAGAATTGCATGATCACATACGCCGCATACAGAAACATA 2244

Db       2161 GAAATAATGGGAGTCTTCAATAGGAACTGCATGATAACATATGCCGCGTATCGAAACATT 2220

Qy       2245 TTCCCCATTTGGGCGTTGGGAGAATACCAATCCCAAGTATT 2285

Db       2221 TTTCCGATATGGGCTTTGGGGGAGTACCGTTGTCAGGTATT 2261


RESULT 10
BT001118
LOCUS       BT001118                2280 bp    mRNA    linear   PLN 08-NOV-2002
DEFINITION  Arabidopsis thaliana At2g07050/T4E14.16 mRNA, complete cds.
ACCESSION   BT001118
VERSION     BT001118.1  GI:24796993
KEYWORDS    FLI_CDNA.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 2280)
  AUTHORS   Shinn,P., Chen,H., Cheuk,R., Kim,C.J., Bowser,L., Carninci,P.,
            Chan,M.M., Chang,C.H., Dale,J.M., Hayashizaki,Y., Hsuan,V.W.,
            Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J.,
            Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M., Nguyen,M.,
            Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M., Southwick,A.,
            Tang,C.C., Toriumi,M., Wallender,E.K., Wong,C., Wu,H.C., Yamada,K.,
            Yu,G., Yuan,S., Shinozaki,K., Davis,R.W., Theologis,A. and
            Ecker,J.R.
  TITLE     Arabidopsis ORF clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2280)
  AUTHORS   Shinn,P., Chen,H., Cheuk,R., Kim,C.J., Bowser,L., Carninci,P.,
            Chan,M.M., Chang,C.H., Dale,J.M., Hayashizaki,Y., Hsuan,V.W.,
            Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J.,
            Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M., Nguyen,M.,
            Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M., Southwick,A.,
            Tang,C.C., Toriumi,M., Wallender,E.K., Wong,C., Wu,H.C., Yamada,K.,
            Yu,G., Yuan,S., Shinozaki,K., Davis,R.W., Theologis,A. and
            Ecker,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-2002) Salk Institute Genomic Analysis Laboratory
            (SIGnAL), Plant Biology Laboratory, The Salk Institute for
            Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
            USA
FEATURES             Location/Qualifiers
     source          1. .2280
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /chromosome="2"
                     /clone="U21897"
                     /ecotype="Columbia"
     CDS             1. .2280
                     /note="cycloartenol synthase"
                     /codon_start=1
                     /product="At2g07050/T4E14.16"
                     /protein_id="AAN64509.1"
                     /db_xref="GI:24796994"
                     /translation="MWKLKIAEGGSPWLRTTNNHVGRQFWEFDPNLGTPEDLAAVEEA
                     RKSFSDNRFVQKHSADLLMRLQFSRENLISPVLPQVKIEDTDDVTEEMVETTLKRGLD
                     FYSTIQAHDGHWPGDYGGPMFLLPGLIITLSITGALNTVLSEQHKQEMRRYLYNHQNE
                     DGGWGLHIEGPSTMFGSVLNYVTLRLLGEGPNDGDGDMEKGRDWILNHGGATNITSWG
                     KMWLSVLGAFEWSGNNPLPPEIWLLPYFLPIHPGRMWCHCRMVYLPMSYLYGKRFVGP
                     ITSTVLSLRKELFTVPYHEVNWNEARNLCAKEDLYYPHPLVQDILWASLHKIVEPVLM
                     RWPGANLREKAIRTAIEHIHYEDENTRYICIGPVNKVLNMLCCWVEDPNSEAFKLHLP
                     RIHDFLWLAEDGMKMQGYNGSQLWDTGFAIQAILATNLVEEYGPVLEKAHSFVKNSQV
                     LEDCPGDLNYWYRHISKGAWPFSTADHGWPISDCTAEGLKAALLLSKVPKAIVGEPID
                     AKRLYEAVNVIISLQNADGGLATYELTRSYPWLELINPAETFGDIVIDYPYVECTSAA
                     IQALISFRKLYPGHRKKEVDECIEKAVKFIESIQAADGSWYGSWAVCFTYGTWFGVKG
                     LVAVGKTLKNSPHVAKACEFLLSKQQPSGGWGESYLSCQDKVYSNLDGNRSHVVNTAW
                     AMLALIGAGQAEVDRKPLHRAARYLINAQMENGDFPQQEIMGVFNRNCMITYAAYRNI
                     FPIWALGEYRCQVLLQQGE"
ORIGIN

  Query Match           54.9%;  Score 1405;  DB 8;  Length 2280;
  Best Local Similarity 76.3%;  Pred. No. 0;
  Matches 1726;  Conservative    0;  Mismatches  535;  Indels    0;  Gaps    0;

Qy         25 ATGTGGAAGCTCAAGTTCGCCGAAGGAGGGAATCCATGGCTTCGGACATTGAACAATCAC 84

Db          1 ATGTGGAAACTGAAGATCGCGGAAGGAGGTAGTCCATGGCTTAGAACCACCAATAATCAC 60

Qy         85 GTTGGAAGACAGGTGTGGGAGTTCGATCCTAAGCTTGGATCGCCGCAAGATCTCCTCGAG 144

Db         61 GTCGGAAGACAGTTTTGGGAGTTCGATCCGAATCTCGGTACTCCTGAGGATCTCGCCGCC 120

Qy        145 ATTGAGAAAGCTCGCCAGAATTTTCACGATAACCGCTTTACCCACAAACACAGCGCTGAT 204

Db        121 GTCGAAGAAGCTAGGAAGTCTTTTTCGGATAATCGATTCGTGCAGAAACATAGCGCCGAT 180

Qy        205 CTACTTATGCGGATGCAGTTCGCAAGAGAGAACCCAACACGTGAAGTCTTGCCCAAAGTC 264

Db        181 CTGCTTATGCGCCTTCAGTTTTCAAGAGAAAATTTGATTAGCCCAGTTTTACCTCAAGTC 240

Qy        265 GGAGTTAAGGATATTGAGGATGTGACCCAAGAGATTGTGACAAAAACATTAAGAAGGGCC 324

Db        241 AAAATCGAAGACACTGATGATGTTACAGAGGAGATGGTGGAAACCACGTTAAAGAGGGGT 300

Qy        325 GTAAGTTTCCATTCAACTCTCCAGTGCCATGACGGACACTGGCCGGGAGATTATGGAGGT 384

Db        301 CTAGATTTCTATTCAACTATACAGGCACACGACGGGCACTGGCCAGGTGATTATGGTGGT 360

Qy        385 CCCATGTTTCTGATGCCTGGCTTGGTAATTACTCTGTCTATCACTGGGGCGTTGAATACA 444

Db        361 CCTATGTTTCTTCTCCCAGGACTGATAATTACACTCTCCATAACTGGAGCACTGAATACA 420

Qy        445 GTCTTAACTGAAGAACATAGAAAGGAAATATGCCGTTACCTCTATAATCATCAAAACAAG 504

Db        421 GTATTGTCGGAACAACATAAACAAGAAATGCGCCGTTATCTCTATAATCACCAGAATGAG 480

Qy        505 GATGGTGGGTGGGGTTTGCATATTGAAGGTCCAAGCACCATGTTTGGCTCTGTCTTGAGT 564

Db        481 GACGGAGGTTGGGGTTTACATATTGAGGGCCCTAGCACCATGTTTGGGTCTGTGTTGAAC 540

Qy        565 TATATTACTCTGAGATTGCTAGGTGAGGGGCCTAATGATGGACAAGGGGAAATGGAGAAG 624

Db        541 TATGTTACTCTAAGGTTGCTTGGAGAAGGACCTAACGATGGAGATGGAGATATGGAGAAA 600

Qy        625 GCACGTGACTGGATTCTAGGGCATGGTGGTGCTACTTATATAACGTCATGGGGGAAGATG 684

Db        601 GGACGAGACTGGATACTAAATCATGGTGGTGCTACCAATATTACATCTTGGGGGAAAATG 660

Qy        685 TGGCTTTCAGTACTTGGAGTGTATGAATGGTCTGGAAATAATCCCCTGCCCCCTGAGATA 744

Db        661 TGGCTATCGGTACTTGGAGCTTTTGAATGGTCCGGAAATAACCCACTGCCACCTGAGATA 720

Qy        745 TGGCTCCTTCCATACATGCTTCCATTTCATCCAGGAAGGATGTGGTGTCACTGCCGGATG 804

Db        721 TGGCTTCTCCCATATTTCCTGCCAATACATCCAGGAAGGATGTGGTGCCATTGTCGAATG 780

Qy        805 GTCTATTTGCCGATGTCCTACTTATATGGCAAGAGGTTTGTTGGTCCAATCTCACCAACA 864

Db        781 GTGTACTTGCCGATGTCGTATTTGTATGGAAAAAGGTTTGTGGGTCCCATAACGTCCACT 840
```

The Scripps Research Institute (US) ; Syngenta Participations AG (CH)

FEATURES Location/Qualifiers
source 1. .2280
/organism="Arabidopsis thaliana"
/mol_type="unassigned DNA"
/db_xref="taxon:3702"

ORIGIN

Query Match 54.9%; Score 1405; DB 6; Length 2280;
Best Local Similarity 76.3%; Pred. No. 0;
Matches 1726; Conservative 0; Mismatches 535; Indels 0; Gaps 0;

```
Qy          25 ATGTGGAAGCTCAAGTTCGCCGAAGGAGGGAATCCATGGCTTCGGACATTGAACAATCAC 84
               ||||||||  ||  ||| ||||  |||||||| | |||||||||| | ||     || ||||||
Db           1 ATGTGGAAACTGAAGATCGCGGAAGGAGGTAGTCCATGGCTTAGAACCACCAATAATCAC 60

Qy          85 GTTGGAAGACAGGTGTGGGAGTTCGATCCTAAGCTTGGATCGCCGCAAGATCTCCTCGAG 144
               || |||||||||  | ||||||||||||||  || ||  ||   | ||   | ||||||  ||
Db          61 GTCGGAAGACAGTTTTGGGAGTTCGATCCGAATCTCGGTACTCCTGAGGATCTCGCCGCC 120

Qy         145 ATTGAGAAAGCTCGCCAGAATTTTCACGATAACCGCTTTACCCACAAACACAGCGCTGAT 204
               | ||  |||||  |  ||  ||||   |||||  || ||     || |||||  ||||| |||
Db         121 GTCGAAGAAGCTAGGAAGTCTTTTTCGGATAATCGATTCGTGCAGAAACATAGCGCCGAT 180

Qy         205 CTACTTATGCGGATGCAGTTCGCAAGAGAGAACCCAACACGTGAAGTCTTGCCCAAAGTC 264
               || |||||||||  | |||||  ||||||| ||     |   ||| || ||  |||||
Db         181 CTGCTTATGCGCCTTCAGTTTTCAAGAGAAAATTTGATTAGCCCAGTTTTACCTCAAGTC 240

Qy         265 GGAGTTAAGGATATTGAGGATGTGACCCAAGAGATTGTGACAAAAACATTAAGAAGGGCC 324
                 | |  | ||  | |||  |||||  ||  | |||||  |||  ||  ||  ||||  ||||
Db         241 AAAATCGAAGACACTGATGATGTTACAGAGGAGATGGTGGAAACCACGTTAAAGAGGGGT 300

Qy         325 GTAAGTTTCCATTCAACTCTCCAGTGCCATGACGGACACTGGCCGGGAGATTATGGAGGT 384
               ||  |||| |||||||| | |||    || |||||  |||||||| || ||||||| |||
Db         301 CTAGATTTCTATTCAACTATACAGGCACACGACGGGCACTGGCCAGGTGATTATGGTGGT 360

Qy         385 CCCATGTTTCTGATGCCTGGCTTGGTAATTACTCTGTCTATCACTGGGGCGTTGAATACA 444
               ||  ||||||||  | ||  ||  || |||||||  || ||  || |||||  ||  ||||||||
Db         361 CCTATGTTTCTTCTCCCAGGACTGATAATTACACTCTCCATAACTGGAGCACTGAATACA 420

Qy         445 GTCTTAACTGAAGAACATAGAAAGGAAATATGCCGTTACCTCTATAATCATCAAAACAAG 504
               ||  ||  | |||  ||||||  | |  |||||  |||||||  ||||||||||| ||  ||  ||
Db         421 GTATTGTCGGAACAACATAAACAAGAAATGCGCCGTTATCTCTATAATCACCAGAATGAG 480

Qy         505 GATGGTGGGTGGGGTTTGCATATTGAAGGTCCAAGCACCATGTTTGGCTCTGTCTTGAGT 564
               || || ||  ||||||||  ||||||||  || ||  ||||||||||||||  |||||  ||||
Db         481 GACGGAGGTTGGGGTTTACATATTGAGGGCCCTAGCACCATGTTTGGGTCTGTGTTGAAC 540

Qy         565 TATATTACTCTGAGATTGCTAGGTGAGGGGCCTAATGATGGACAAGGGGAAATGGAGAAG 624
               |||  |||||||  ||  |||||  ||  ||  ||  |||||  ||||||  |  ||  ||  ||||||||
Db         541 TATGTTACTCTAAGGTTGCTTGGAGAAGGACCTAACGATGGAGATGGAGATATGGAGAAA 600

Qy         625 GCACGTGACTGGATTCTAGGGCATGGTGGTGCTACTTATATAACGTCATGGGGGAAGATG 684
               | |||  ||||||||  |||   ||||||||||||||  ||||  ||  ||  ||||||||  |||
Db         601 GGACGAGACTGGATACTAAATCATGGTGGTGCTACCAATATTACATCTTGGGGGAAAATG 660

Qy         685 TGGCTTTCAGTACTTGGAGTGTATGAATGGTCTGGAAATAATCCCCTGCCCCCTGAGATA 744
               |||||  ||  ||||||||||  |  ||||||||||  ||||||||  ||  |||||  ||||||||||
Db         661 TGGCTATCGGTACTTGGAGCTTTTGAATGGTCCGGAAATAACCCACTGCCACCTGAGATA 720

Qy         745 TGGCTCCTTCCATACATGCTTCCATTTCATCCAGGAAGGATGTGGTGTCACTGCCGGATG 804
               |||||  ||  |||||  |  ||  |||  | ||||||||||||||||||||  ||  ||  |||
Db         721 TGGCTTCTCCCATATTTCCTGCCAATACATCCAGGAAGGATGTGGTGCCATTGTCGAATG 780

Qy         805 GTCTATTTGCCGATGTCCTACTTATATGGCAAGAGGTTTGTTGGTCCAATCTCACCAACA 864
               ||  ||  ||||||||||||  ||  ||  |||||  ||  ||||||||  |||||  ||  |  | ||
Db         781 GTGTACTTGCCGATGTCGTATTTGTATGGAAAAAGGTTTGTGGGTCCCATAACGTCCACT 840

Qy         865 GTATTATCTTTGAGAAAAGAGCTTTATACAGTACCATACCATGATATAGATTGGGATCAG 924
               ||  |||||  |||||||  ||||||||  ||||||||||  |||||  |  | |||  ||  |
Db         841 GTTTTATCACTGAGAAAGGAGCTTTTCACAGTACCATATCATGAAGTCAACTGGAATGAA 900

Qy         925 GCTCGCAATTTGTGTGCAAAGGAAGATTTGTACTATCCTCACCCACTTGTACAGGATATT 984
               ||  |||||  | ||  |||||||| |||||  |||||  ||  ||  ||||||||  ||  ||||||
Db         901 GCACGCAACCTTTGCGCAAAGGAGGATTTATACTACCCACATCCACTTGTGCAAGATATT 960

Qy         985 CTTTGGGCATCTCTACACAAGTTCCTTGAGCCTATTCTGATGCATTGGCCTGGAAAAAGA 1044
               ||||||||||  ||  ||  |||  |  ||||||||  |||||||||  ||||||||    ||
Db         961 CTTTGGGCATCACTTCATAAGATTGTTGAGCCTGTTCTGATGCGATGGCCTGGTGCAAAT 1020

Qy        1045 TTGAGGGAAAAGGCTATTATTTCTGCATTGGAGCATATACATTACGAAGATGAGAATACT 1104
               |||||  ||||||||||  |    | |||  | ||  |||||  |||||  ||||||||||||||
Db        1021 TTGAGAGAAAAGGCTATAAGAACCGCAATAGAACATATTCATTATGAAGATGAGAATACT 1080

Qy        1105 CGATATATTTGCATAGGTCCTGTAAATAAGGTGTTAAATATGCTTTGCTGTTGGGTGGAA 1164
               | || ||  ||||||||||  ||  ||  |||||  |||||||||||||||||||||||  |||
Db        1081 AGGTACATCTGCATAGGTCCCGTGAACAAGGTATTAAATATGCTTTGCTGTTGGGTAGAA 1140

Qy        1165 GATCCAAATTCTGAGGCCTTCAAGTTGCATCTTCCCAGGATTTATGATTATCTATGGATT 1224
               ||  |||||  ||  |||||  ||||||||||  ||  ||  ||  ||||  |  |||  |||  |
Db        1141 GACCCAAACTCAGAGGCTTTCAAGTTGCACCTACCAAGAATCCATGACTTTCTCTGGTTA 1200

Qy        1225 GCAGAAGATGGCATGAAAATGCAGGGCTACAATGGAAGTCAACTATGGGACACTGCTTTT 1284
               ||  |||||||  |||||  |||||||  ||  ||  |||||  ||  ||||||||  ||  |  ||||
Db        1201 GCTGAAGATGGAATGAAGATGCAGGGTTATAACGGAAGCCAGCTATGGGATACAGGTTTT 1260

Qy        1285 GCTGTCCAAGCAATTATTGCATCTAACCTCATTGAAGAATTTGGTCCAACTATAAGAAAA 1344
               |||  |  |||||  |||  |  |||  ||||||||  |  |||||||  |||  ||    |  ||||
Db        1261 GCTATTCAAGCGATTTTGGCAACTAACCTCGTCGAAGAATATGGGCCCGTTTTGGAAAAA 1320

Qy        1345 GCTCATACCTATATTAAGAATTCACAGGTTTTAGAAGATTGTCCAGGTGATCTTAATAAA 1404
               ||  |||  |  |  |  ||||||||  |||||  ||||||||  ||  ||  ||  |||||  |||  |
Db        1321 GCACATTCATTTGTCAAGAATTCCCAGGTGTTAGAAGACTGCCCTGGAGATCTGAATTAC 1380

Qy        1405 TGGTACCGTCACATTTCAAAAGGTGCTTGGCCTTTTTCAACTGGAGATCATGGATGGCCA 1464
               |||||  ||  ||||||||  |||||  ||||||||||  |||||||  |||||  ||  |||||
Db        1381 TGGTATCGCCACATTTCTAAAGGGGCTTGGCCTTTCTCAACTGCAGATCACGGTTGGCCC 1440

Qy        1465 ATTTCTGACTGCACAGCTGAAGGACTGAAAGCTGTTCTATTACTATCCAAAATTGCACCA 1524
               ||  ||||||||||  ||  ||||||||||||||||  |||  ||  |||||||||  ||  |
Db        1441 ATCTCTGACTGCACCGCAGAAGGACTGAAAGCTGCTCTTTTGCTATCCAAAGTTCCCAAG 1500

Qy        1525 GAAATAGTTGGTGAGCCAATAGACGTGAAGCGATTATATGATTCTGTAAATGTCATTCTC 1584
               |  ||  ||||||||  ||||||||  |  ||  ||  ||||||||  ||||  |||||  ||  |
Db        1501 GCGATTGTTGGTGAACCAATAGATGCAAAACGGTTATATGAAGCTGTTAATGTTATCATT 1560

Qy        1585 TCACTACAGAATGAAGATGGTGGTTTTGCAACATATGAGCTTAAACGATCTTATAATTGG 1644
               ||  |||||||  ||||||  ||  |  ||||||||||||||  |  |||  ||  ||||
Db        1561 TCTTTACAGAATGCAGATGGAGGCCTCGCAACATATGAGCTCACCAGGTCATACCCTTGG 1620

Qy        1645 TTGGAGATAATCAATCCTGCTGAAACTTTTGGTGACATCGTTATTGATTATCCTTATGTG 1704
               ||  |||  ||||||||  ||  ||  |||||  |||||  ||  ||  ||||||||||||||||  |||
Db        1621 TTAGAGCTAATCAACCCAGCAGAAACCTTTGGCGATATTGTTATTGATTATCCTTACGTG 1680

Qy        1705 GAATGTACATCAGCAGCGATTCAAGCTTTGGCATCATTTAGGAAATTATATCCTGGGCAT 1764
               |||||||||||||||  ||  ||  |||||||||  |||||||  |  ||  |  ||||||||  |||
Db        1681 GAATGTACATCAGCTGCTATCCAAGCTTTGATATCATTTCGAAAGCTGTATCCTGGTCAT 1740

Qy        1765 CGCCGAGAAGAAATACAACATTGTATCGATAAAGCCACTACCTTCATTGAAAAAATACAA 1824
               ||    | |||  ||  |  |  ||  ||  ||  ||  ||  |||||||||  ||  |||
Db        1741 CGAAAGAAGGAAGTAGATGAGTGCATTGAGAAGGCGGTTAAGTTCATTGAATCCATTCAA 1800

Qy        1825 GCTTCAGATGGATCATGGTATGGTTCTTGGGGAGTTTGCTTCACTTACGGTGCTTGGTTT 1884
               ||  |||||||  ||||||||||  ||  ||||  ||||||||||  ||  |||  |  ||||||
Db        1801 GCAGCAGATGGCTCATGGTATGGATCATGGGCTGTTTGCTTCACGTATGGTACGTGGTTT 1860

Qy        1885 GGGGTAAAAGGTCTGATTGCTGCTGGAAGGAGTTTCAGTAATTGCTCAAGCATCCGTAAA 1944
               ||  ||  |||||  |||  |  ||||  |||||  |  ||  |  ||  |  ||  |  ||||
Db        1861 GGAGTGAAAGGGCTGGTAGCTGTTGGAAAAACATTGAAAAACTCTCCACATGTTGCTAAA 1920

Qy        1945 GCTTGTGAATTTCTGCTGTCCAAGCAGCTTCCTTCTGGTGGCTGGGGAGAGAGTTATCTG 2004
               ||||||||||||||  ||||  ||  ||  |  |||||  ||  ||||||||||  ||  |||||
Db        1921 GCTTGTGAATTTCTATTGTCGAAACAACAACCTTCGGGCGGCTGGGGAGAAAGCTATCTT 1980

Qy        2005 TCCTGTCAAAACAAGGTGTATTCAAATCTGGAAGGCAACAGGTCTCATGTGGTCAACACT 2064
```

```
Qy       421 TCTATCACTGGGGCGTTGAATACAGTCTTAACTGAAGAACATAGAAAGGAAATATGCCGT 480
Db       579 TCCATAACTGGAGCACTGAATACAGTATTGTCGGAACAACATAAACAAGAAATGCGCCGT 638

Qy       481 TACCTCTATAATCATCAAAACAAGGATGGTGGGTGGGGTTTGCATATTGAAGGTCCAAGC 540
Db       639 TATCTCTATAATCACCAGAATGAGGACGGAGGTTGGGGTTTACATATTGAGGGCCCTAGC 698

Qy       541 ACCATGTTTGGCTCTGTCTTGAGTTATATTACTCTGAGATTGCTAGGTGAGGGGCCTAAT 600
Db       699 ACCATGTTTGGGTCTGTGTTGAACTATGTTACTCTAAGGTTGCTTGGAGAAGGACCTAAC 758

Qy       601 GATGGACAAGGGGAAATGGAGAAGGCACGTGACTGGATTCTAGGGCATGGTGGTGCTACT 660
Db       759 GATGGAGATGGAGATATGGAGAAAGGACGAGACTGGATACTAAATCATGGTGGTGCTACC 818

Qy       661 TATATAACGTCATGGGGGAAGATGTGGCTTTCAGTACTTGGAGTGTATGAATGGTCTGGA 720
Db       819 AATATTACATCTTGGGGGAAAATGTGGCTATCGGTACTTGGAGCTTTTGAATGGTCCGGA 878

Qy       721 AATAATCCCCTGCCCCCTGAGATATGGCTCCTTCCATACATGCTTCCATTTCATCCAGGA 780
Db       879 AATAACCCACTGCCACCTGAGATATGGCTTCTCCCATATTTCCTGCCAATACATCCAGGA 938

Qy       781 AGGATGTGGTGTCACTGCCGGATGGTCTATTTGCCGATGTCCTACTTATATGGCAAGAGG 840
Db       939 AGGATGTGGTGCCATTGTCGAATGGTGTACTTGCCGATGTCGTATTTGTATGGAAAAAGG 998

Qy       841 TTTGTTGGTCCAATCTCACCAACAGTATTATCTTTGAGAAAAGAGCTTTATACAGTACCA 900
Db       999 TTTGTGGGTCCCATAACGTCCACTGTTTTATCACTGAGAAAGGAGCTTTTCACAGTACCA 1058

Qy       901 TACCATGATATAGATTGGGATCAGGCTCGCAATTTGTGTGCAAAGGAAGATTTGTACTAT 960
Db      1059 TATCATGAAGTCAACTGGAATGAAGCACGCAACCTTTGCGCAAAGGAGGATTTATACTAC 1118

Qy       961 CCTCACCCACTTGTACAGGATATTCTTTGGGCATCTCTACACAAGTTCCTTGAGCCTATT 1020
Db      1119 CCACATCCACTTGTGCAAGATATTCTTTGGGCATCACTTCATAAGATTGTTGAGCCTGTT 1178

Qy      1021 CTGATGCATTGGCCTGGAAAAAGATTGAGGGAAAAGGCTATTATTTCTGCATTGGAGCAT 1080
Db      1179 CTGATGCGATGGCCTGGTGCAAATTTGAGAGAAAAGGCTATAAGAACCGCAATAGAACAT 1238

Qy      1081 ATACATTACGAAGATGAGAATACTCGATATATTTGCATAGGTCCTGTAAATAAGGTGTTA 1140
Db      1239 ATTCATTATGAAGATGAGAATACTAGGTACATCTGCATAGGTCCCGTGAACAAGGTATTA 1298

Qy      1141 AATATGCTTTGCTGTTGGGTGGAAGATCCAAATTCTGAGGCCTTCAAGTTGCATCTTCCC 1200
Db      1299 AATATGCTTTGCTGTTGGGTAGAAGACCCAAACTCAGAGGCTTTCAAGTTGCACCTACCA 1358

Qy      1201 AGGATTTATGATTATCTATGGATTGCAGAAGATGGCATGAAAATGCAGGGCTACAATGGA 1260
Db      1359 AGAATCCATGACTTTCTCTGGTTAGCTGAAGATGGAATGAAGATGCAGGGTTATAACGGA 1418

Qy      1261 AGTCAACTATGGGACACTGCTTTTGCTGTCCAAGCAATTATTGCATCTAACCTCATTGAA 1320
Db      1419 AGCCAGCTATGGGATACAGGTTTTGCTATTCAAGCGATTTTGGCAACTAACCTCGTCGAA 1478

Qy      1321 GAATTTGGTCCAACTATAAGAAAAGCTCATACCTATATTAAGAATTCACAGGTTTTAGAA 1380
Db      1479 GAATATGGGCCCGTTTTGGAAAAAGCACATTCATTTGTCAAGAATTCCCAGGTGTTAGAA 1538

Qy      1381 GATTGTCCAGGTGATCTTAATAAATGGTACCGTCACATTTCAAAAGGTGCTTGGCCTTTT 1440
Db      1539 GACTGCCCTGGAGATCTGAATTACTGGTATCGCCACATTTCTAAAGGGGCTTGGCCTTTC 1598

Qy      1441 TCAACTGGAGATCATGGATGGCCAATTTCTGACTGCACAGCTGAAGGACTGAAAGCTGTT 1500
Db      1599 TCAACTGCAGATCACGGTTGGCCCATCTCTGACTGCACCGCAGAAGGACTGAAAGCTGCT 1658

Qy      1501 CTATTACTATCCAAAATTGCACCAGAAATAGTTGGTGAGCCAATAGACGTGAAGCGATTA 1560

Db      1659 CTTTTGCTATCCAAAGTTCCCAAGGCGATTGTTGGTGAACCAATAGATGCAAAACGGTTA 1718

Qy      1561 TATGATTCTGTAAATGTCATTCTCTCACTACAGAATGAAGATGGTGGTTTTGCAACATAT 1620
Db      1719 TATGAAGCTGTTAATGTTATCATTTCTTTACAGAATGCAGATGGAGGCCTCGCAACATAT 1778

Qy      1621 GAGCTTAAACGATCTTATAATTGGTTGGAGATAATCAATCCTGCTGAAACTTTTGGTGAC 1680
Db      1779 GAGCTCACCAGGTCATACCCTTGGTTAGAGCTAATCAACCCAGCAGAAACCTTTGGCGAT 1838

Qy      1681 ATCGTTATTGATTATCCTTATGTGGAATGTACATCAGCAGCGATTCAAGCTTTGGCATCA 1740
Db      1839 ATTGTTATTGATTATCCTTACGTGGAATGTACATCAGCTGCTATCCAAGCTTTGATATCA 1898

Qy      1741 TTTAGGAAATTATATCCTGGGCATCGCCGAGAAGAAATACAACATTGTATCGATAAAGCC 1800
Db      1899 TTTCGAAAGCTGTATCCTGGTCATCGAAAGAAGGAAGTAGATGAGTGCATTGAGAAGGCG 1958

Qy      1801 ACTACCTTCATTGAAAAAATACAAGCTTCAGATGGATCATGGTATGGTTCTTGGGGAGTT 1860
Db      1959 GTTAAGTTCATTGAATCCATTCAAGCAGCAGATGGCTCATGGTATGGATCATGGGCTGTT 2018

Qy      1861 TGCTTCACTTACGGTGCTTGGTTTGGGGTAAAAGGTCTGATTGCTGCTGGAAGGAGTTTC 1920
Db      2019 TGCTTCACGTATGGTACGTGGTTTGGAGTGAAAGGGCTGGTAGCTGTTGGAAAAACATTG 2078

Qy      1921 AGTAATTGCTCAAGCATCCGTAAAGCTTGTGAATTTCTGCTGTCCAAGCAGCTTCCTTCT 1980
Db      2079 AAAAACTCTCCACATGTTGCTAAAGCTTGTGAATTTCTATTGTCGAAACAACAACCTTCG 2138

Qy      1981 GGTGGCTGGGGAGAGAGTTATCTGTCCTGTCAAAACAAGGTGTATTCAAATCTGGAAGGC 2040
Db      2139 GGCGGCTGGGGAGAAAGCTATCTTTCATGTCAAGACAAGGTCTATTCAAACCTTGATGGC 2198

Qy      2041 AACAGGTCTCATGTGGTCAACACTGGGTGGGCTATGTTGGCTCTCATTGATGCTGGACAG 2100
Db      2199 AACAGATCTCACGTCGTGAATACAGCATGGGCTATGCTCGCACTCATTGGTGCTGGGCAA 2258

Qy      2101 GCTAAGAGAGATTCGCAACCACTGCACCGGGCAGCTGCATACTTGATAAATTCCCAATTG 2160
Db      2259 GCTGAGGTAGACCGGAAACCACTACACCGGGCTGCAAGATACTTGATTAATGCTCAAATG 2318

Qy      2161 GAGGACGGTGACTTTCCGCAGCAGGAAATAATGGGAGTCTTCAACAAGAATTGCATGATC 2220
Db      2319 GAGAATGGTGATTTTCCACAACAGGAAATAATGGGAGTCTTCAATAGGAACTGCATGATA 2378

Qy      2221 ACATACGCCGCATACAGAAACATATTCCCCATTTGGGCGTTGGGAGAATACCAATCCCAA 2280
Db      2379 ACATATGCCGCGTATCGAAACATTTTTCCGATATGGGCTTTGGGGGAGTACCGTTGTCAG 2438

Qy      2281 GTATT 2285
Db      2439 GTATT 2443
```

```
RESULT 9
AX506138
LOCUS       AX506138               2280 bp    DNA     linear   PAT 27-SEP-2002
DEFINITION  Sequence 833 from Patent WO0216655.
ACCESSION   AX506138
VERSION     AX506138.1  GI:23387375
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Harper,J.F., Kreps,J., Wang,X. and Zhu,T.
  TITLE     Stress-regulated genes of plants, transgenic plants containing
            same, and methods of use
  JOURNAL   Patent: WO 0216655-A 833 28-FEB-2002;
```

```
Qy       1921 AGTAATTGCTCAAGCATCCGTAAAGCTTGTGAATTTCTGCTGTCCAAGCAGCTTCCTTCT 1980
Db       1950 AAAAACTCTCCACATGTTGCTAAAGCTTGTGAATTTCTATTGTCGAAACAACAACCTTCG 2009

Qy       1981 GGTGGCTGGGGAGAGAGTTATCTGTCCTGTCAAAACAAGGTGTATTCAAATCTGGAAGGC 2040
Db       2010 GGCGGCTGGGGAGAAAGCTATCTTTCATGTCAAGACAAGGTCTATTCAAACCTTGATGGC 2069

Qy       2041 AACAGGTCTCATGTGGTCAACACTGGGTGGGCTATGTTGGCTCTCATTGATGCTGGACAG 2100
Db       2070 AACAGATCTCACGTCGTGAATACAGCATGGGCTATGCTCGCACTCATTGGTGCTGGGCAA 2129

Qy       2101 GCTAAGAGAGATTCGCAACCACTGCACCGGGCAGCTGCATACTTGATAAATTCCCAATTG 2160
Db       2130 GCTGAGGTAGACCGGAAACCACTACACCGGGCTGCAAGATACTTGATTAATGCTCAAATG 2189

Qy       2161 GAGGACGGTGACTTTCCGCAGCAGGAAATAATGGGAGTCTTCAACAAGAATTGCATGATC 2220
Db       2190 GAGAATGGTGATTTTCCACAACAGGAAATAATGGGAGTCTTCAATAGGAACTGCATGATA 2249

Qy       2221 ACATACGCCGCATACAGAAACATATTCCCCATTTGGGCGTTGGGAGAATACCAATCCCAA 2280
Db       2250 ACATATGCCGCGTATCGAAACATTTTTCCGATATGGGCGTTGGGGGAGTACCGTTGTCAG 2309

Qy       2281 GTATT 2285
Db       2310 GTATT 2314
```

RESULT 8
AY094394

```
LOCUS       AY094394                2654 bp    mRNA    linear   PLN 05-MAY-2002
DEFINITION  Arabidopsis thaliana At2g07050/T4E14.16 mRNA, complete cds.
ACCESSION   AY094394
VERSION     AY094394.1  GI:20453055
KEYWORDS    FLI_CDNA.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 2654)
  AUTHORS   Shinn,P., Chen,H., Cheuk,R., Kim,C.J., Meyers,M.C., Banh,J.,
            Bowser,L., Carninci,P., Chang,E., Dale,J.M., Goldsmith,A.D.,
            Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G.,
            Kawai,J., Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M.,
            Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L., Sakurai,T.,
            Satou,M., Seki,M., Southwick,A., Tang,C.C., Toriumi,M., Wu,H.C.,
            Yamada,K., Yamamura,Y., Yu,G., Yu,S., Shinozaki,K., Davis,R.W.,
            Theologis,A. and Ecker,J.R.
  TITLE     Arabidopsis cDNA clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2654)
  AUTHORS   Shinn,P., Chen,H., Cheuk,R., Kim,C.J., Meyers,M.C., Banh,J.,
            Bowser,L., Carninci,P., Chang,E., Dale,J.M., Goldsmith,A.D.,
            Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G.,
            Kawai,J., Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M.,
            Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L., Sakurai,T.,
            Satou,M., Seki,M., Southwick,A., Tang,C.C., Toriumi,M., Wu,H.C.,
            Yamada,K., Yamamura,Y., Yu,G., Yu,S., Shinozaki,K., Davis,R.W.,
            Theologis,A. and Ecker,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-APR-2002) Salk Institute Genomic Analysis Laboratory
            (SIGnAL), Plant Biology Laboratory, The Salk Institute for
            Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
            USA
COMMENT     RIKEN Genomic Sciences Center (GSC) members carried out the
            collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
            Arabidopsis Full-Length cDNA') :  Seki,M., Narusaka,M., Ishida,J.,
            Satou,M., Kamiya,A., Sakurai,T., Carninci,P.,  Kawai,J.,
            Hayashizaki,Y. and Shinozaki,K.

            The Salk, Stanford, PGEC (SSP) Consortium members carried out the
            sequencing and annotation of the RAFL cDNAs: Shinn,P., Chen,H.,
            Cheuk,R., Kim,C.J.,  Meyers,M.C., Banh,J. Bowser,L., Chang,E.,
            Dale,J.M., Goldsmith,A.D., Jones,T., Karlin-Neumann,G., Lam,B.,
            Lee,J.M., Lin,J., Miranda,M., Nguyen,M., Onodera,C.S., Palm,C.J.,
            Quach,H.L., Southwick,A., Tang,C.C., Toriumi,M., Wu,H.C.,
            Yamada,K., Yamamura, Y., Yu,G., Yu,S., Davis,R.W.,  Theologis,A.,
            and Ecker,J.R.

            Shinn,P. (SSP/Salk) and Seki,M. (RIKEN GSC) contributed equally to
            this work.  Shinozaki,K. (RIKEN GSC) and Ecker,J.R. (SSP/Salk)
            contributed equally to this work as PIs.
FEATURES             Location/Qualifiers
     source          1. .2654
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /chromosome="2"
                     /clone="RAFL09-12-N05(R21897)"
                     /ecotype="Columbia"
     CDS             183. .2462
                     /note="cycloartenol synthase"
                     /codon_start=1
                     /product="At2g07050/T4E14.16"
                     /protein_id="AAM19773.1"
                     /db_xref="GI:20453056"
                     /translation="MWKLKIAEGGSPWLRTTNNHVGRQFWEFDPNLGTPEDLAAVEEA
                     RKSFSDNRFVQKHSADLLMRLQFSRENLISPVLPQVKIEDTDDVTEEMVETTLKRGLD
                     FYSTIQAHDGHWPGDYGGPMFLLPGLIITLSITGALNTVLSEQHKQEMRRYLYNHQNE
                     DGGWGLHIEGPSTMFGSVLNYVTLRLLGEGPNDGDGDMEKGRDWILNHGGATNITSWG
                     KMWLSVLGAFEWSGNNPLPPEIWLLPYFLPIHPGRMWCHCRMVYLPMSYLYGKRFVGP
                     ITSTVLSLRKELFTVPYHEVNWNEARNLCAKEDLYYPHPLVQDILWASLHKIVEPVLM
                     RWPGANLREKAIRTAIEHIHYEDENTRYICIGPVNKVLNMLCCWVEDPNSEAFKLHLP
                     RIHDFLWLAEDGMKMQGYNGSQLWDTGFAIQAILATNLVEEYGPVLEKAHSFVKNSQV
                     LEDCPGDLNYWYRHISKGAWPFSTADHGWPISDCTAEGLKAALLLSKVPKAIVGEPID
                     AKRLYEAVNVIISLQNADGGLATYELTRSYPWLELINPAETFGDIVIDYPYVECTSAA
                     IQALISFRKLYPGHRKKEVDECIEKAVKFIESIQAADGSWYGSWAVCFTYGTWFGVKG
                     LVAVGKTLKNSPHVAKACEFLLSKQQPSGGWGESYLSCQDKVYSNLDGNRSHVVNTAW
                     AMLALIGAGQAEVDRKPLHRAARYLINAQMENGDFPQQEIMGVFNRNCMITYAAYRNI
                     FPIWALGEYRCQVLLQQGE"
ORIGIN
```

```
  Query Match             55.0%;  Score 1408.2;  DB 8;  Length 2654;
  Best Local Similarity   76.0%;  Pred. No. 0;
  Matches 1737;  Conservative    0;  Mismatches  548;  Indels    0;  Gaps    0;

Qy          1 TTGGCCTCTTGCCAGCAAAACAGAATGTGGAAGCTCAAGTTCGCCGAAGGAGGGAATCCA 60
Db        159 TTTTTCCCTCTCGAGAGAGCGATAATGTGGAAACTGAAGATCGCGGAAGGAGGTAGTCCA 218

Qy         61 TGGCTTCGGACATTGAACAATCACGTTGGAAGACAGGTGTGGGAGTTCGATCCTAAGCTT 120
Db        219 TGGCTTAGAACCACCAATAATCACGTCGGAAGACAGTTTTGGGAGTTCGATCCGAATCTC 278

Qy        121 GGATCGCCGCAAGATCTCCTCGAGATTGAGAAAGCTCGCCAGAATTTTCACGATAACCGC 180
Db        279 GGTACTCCTGAGGATCTCGCCGCCGTCGAAGAAGCTAGGAAGTCTTTTTCGGATAATCGA 338

Qy        181 TTTACCCACAAACACAGCGCTGATCTACTTATGCGGATGCAGTTCGCAAGAGAGAACCCA 240
Db        339 TTCGTGCAGAAACATAGCGCCGATCTGCTTATGCGCCTTCAGTTTCAAGAGAAAATTTG 398

Qy        241 ACACGTGAAGTCTTGCCCAAAGTCGGAGTTAAGGATATTGAGGATGTGACCCAAGAGATT 300
Db        399 ATTAGCCCAGTTTTACCTCAAGTCAAAATCGAAGACACTGATGATGTTACAGAGGAGATG 458

Qy        301 GTGACAAAAACATTAAGAAGGGCCGTAAGTTTCCATTCAACTCTCCAGTGCCATGACGGA 360
Db        459 GTGGAAACCACGTTAAAGAGGGGTCTAGATTTCTATTCAACTATACAGGCACACGACGGG 518

Qy        361 CACTGGCCGGGAGATTATGGAGGTCCCATGTTTCTGATGCCTGGCTTGGTAATTACTCTG 420
Db        519 CACTGGCCAGGTGATTATGGTGGTCCTATGTTTCTTCTCCCAGGACTGATAATTACACTC 578
```

FYSTIQAHDGHWPGDYGGPMFLLPGLIITLSITGALNTVLSEQHKQEMRRYLYNHQNE
DGGWGLHIEGPSTMFGSVLNYVTLRLLGEGPNDGDGDMEKGRDWILNHGGATNITSWG
KMWLSVLGAFEWSGNNPLPPEIWLLPYFLPIHPGRMWCHCRMVYLPMSYLYGKRFVGP
ITSTVLSLRKELFTVPYHEVNWNEARNLCAKEDLYYPHPLVQDILWASLHKIVEPVLM
RWPGANLREKAIRTAIEHIHYEDENTRYICIGPVNKVLNMLCCWVEDPNSEAFKLHLP
RIHDFLWLAEDGMKMQGYNGSQLWDTGFAIQAILATNLVEEYGPVLEKAHSFVKNSQV
LEDCPGDLNYWYRHISKGAWPFSTADHGWPISDCTAEGLKAALLLSKVPKEIVGEPID
AKRLYEAVNVIISLQNADGGLATYELTRSYPWLELINPAETFGDIVIDYPYVECTSAA
IQALISFRKLYPGHRKKEVDECIEKAVKFIESIQAADGSWYGSWAVCFTYGTWFGVKG
LVAVGKTLKNSPHVAKACEFLLSKQQPSGGWGESYLSCQDKVYSNLDGNRSHVVNTAW
AMLALIGAGQAEVDRKPLHRAARYLINAQMENGDFPQQEIMGVFNRNCMITYAAYRNI
FPIWALGEYRCQVLLQQGE"

ORIGIN

Query Match 55.1%; Score 1411.4; DB 8; Length 2561;
Best Local Similarity 76.1%; Pred. No. 0;
Matches 1739; Conservative 0; Mismatches 546; Indels 0; Gaps 0;

```
Qy    1 TTGGCCTCTTGCCAGCAAAACAGAATGTGGAAGCTCAAGTTCGCCGAAGGAGGGAATCCA 60
Db   30 TTTTTCCCTCTCGAGAGAGTGATAATGTGGAAACTGAAGATCGCGGAAGGAGGTAGTCCA 89

Qy   61 TGGCTTCGGACATTGAACAATCACGTTGGAAGACAGGTGTGGGAGTTCGATCCTAAGCTT 120
Db   90 TGGCTTAGAACCACCAATAATCACGTCGGAAGACAGTTTTGGGAGTTCGATCCGAATCTC 149

Qy  121 GGATCGCCGCAAGATCTCCTCGAGATTGAGAAAGCTCGCCAGAATTTTCACGATAACCGC 180
Db  150 GGTACTCCTGAGGATCTCGCCGCCGTCGAAGAAGCTAGGAAGTCTTTTTCGGATAATCGA 209

Qy  181 TTTACCCACAAACACAGCGCTGATCTACTTATGCGGATGCAGTTCGCAAGAGAGAACCCA 240
Db  210 TTCGTGCAGAAACATAGCGCCGATCTGCTTATGCGCCTTCAGTTTTCAAGAGAAAATTTG 269

Qy  241 ACACGTGAAGTCTTGCCCAAAGTCGGAGTTAAGGATATTGAGGATGTGACCCAAGAGATT 300
Db  270 ATTAGCCCAGTTTTACCTCAAGTCAAAATCGAAGACACTGATGATGTTACAGAGGAGATG 329

Qy  301 GTGACAAAAACATTAAGAAGGGCCGTAAGTTTCCATTCAACTCTCCAGTGCCATGACGGA 360
Db  330 GTGGAAACCACGTTAAAGAGGGGTCTAGATTTCTATTCAACTATACAGGCACACGACGGG 389

Qy  361 CACTGGCCGGGAGATTATGGAGGTCCCATGTTTCTGATGCCTGGCTTGGTAATTACTCTG 420
Db  390 CACTGGCCAGGTGATTATGGTGGTCCTATGTTTCTTCTCCCAGGACTGATAATTACACTC 449

Qy  421 TCTATCACTGGGGCGTTGAATACAGTCTTAACTGAAGAACATAGAAAGGAAATATGCCGT 480
Db  450 TCCATAACTGGAGCACTGAATACAGTATTGTCGGAACAACATAAACAAGAAATGCGCCGT 509

Qy  481 TACCTCTATAATCATCAAAACAAGGATGGTGGGTGGGGTTTGCATATTGAAGGTCCAAGC 540
Db  510 TATCTCTATAATCACCAGAATGAGGACGGAGGTTGGGGTTTACATATTGAGGGCCCTAGC 569

Qy  541 ACCATGTTTGGCTCTGTCTTGAGTTATATTACTCTGAGATTGCTAGGTGAGGGGCCTAAT 600
Db  570 ACCATGTTTGGGTCTGTGTTGAACTATGTTACTCTAAGGTTGCTTGGAGAAGGACCTAAC 629

Qy  601 GATGGACAAGGGGAAATGGAGAAGGCACGTGACTGGATTCTAGGGCATGGTGGTGCTACT 660
Db  630 GATGGAGATGGAGATATGGAGAAAGGACGAGACTGGATACTAAATCATGGTGGTGCTACC 689

Qy  661 TATATAACGTCATGGGGGAAGATGTGGCTTTCAGTACTTGGAGTGTATGAATGGTCTGGA 720
Db  690 AATATTACATCTTGGGGGAAAATGTGGCTATCGGTACTTGGAGCTTTTGAATGGTCCGGA 749

Qy  721 AATAATCCCCTGCCCCCTGAGATATGGCTCCTTCCATACATGCTTCCATTTCATCCAGGA 780
Db  750 AATAACCCACTGCCACCTGAGATATGGCTTCTCCCATATTTCCTGCCAATACATCCAGGA 809

Qy  781 AGGATGTGGTGTCACTGCCGGATGGTCTATTTGCCGATGTCCTACTTATATGGCAAGAGG 840
Db  810 AGGATGTGGTGCCATTGTCGAATGGTGTACTTGCCGATGTCGTATTTGTATGGAAAAAGG 869

Qy  841 TTTGTTGGTCCAATCTCACCAACAGTATTATCTTTGAGAAAAGAGCTTTATACAGTACCA 900
Db  870 TTTGTGGGTCCCATAACGTCCACTGTTTTATCACTGAGAAAGGAGCTTTTCACAGTACCA 929

Qy  901 TACCATGATATAGATTGGGATCAGGCTCGCAATTTGTGTGCAAAGGAAGATTTGTACTAT 960
Db  930 TATCATGAAGTCAACTGGAATGAAGCACGCAACCTTTGCGCAAAGGAGGATTTATACTAC 989

Qy  961 CCTCACCCACTTGTACAGGATATTCTTTGGGCATCTCTACACAAGTTCCTTGAGCCTATT 1020
Db  990 CCACATCCACTTGTGCAAGATATTCTTTGGGCATCACTTCATAAGATTGTTGAGCCTGTT 1049

Qy 1021 CTGATGCATTGGCCTGGAAAAAGATTGAGGGAAAAGGCTATTATTTCTGCATTGGAGCAT 1080
Db 1050 CTGATGCGATGGCCTGGTGCAAATTTGAGAGAAAAGGCTATAAGAACCGCAATAGAACAT 1109

Qy 1081 ATACATTACGAAGATGAGAATACTCGATATATTTGCATAGGTCCTGTAAATAAGGTGTTA 1140
Db 1110 ATTCATTATGAAGATGAGAATACTAGGTACATCTGCATAGGTCCCGTGAACAAGGTATTA 1169

Qy 1141 AATATGCTTTGCTGTTGGGTGGAAGATCCAAATTCTGAGGCCTTCAAGTTGCATCTTCCC 1200
Db 1170 AATATGCTTTGCTGTTGGGTAGAAGACCCAAACTCAGAGGCTTTCAAGTTGCACCTACCA 1229

Qy 1201 AGGATTTATGATTATCTATGGATTGCAGAAGATGGCATGAAAATGCAGGGCTACAATGGA 1260
Db 1230 AGAATCCATGACTTTCTCTGGTTAGCTGAAGATGGAATGAAGATGCAGGGTTATAACGGA 1289

Qy 1261 AGTCAACTATGGGACACTGCTTTTGCTGTCCAAGCAATTATTGCATCTAACCTCATTGAA 1320
Db 1290 AGCCAGCTATGGGATACAGGTTTTGCTATTCAAGCGATTTTGGCAACTAACCTCGTCGAA 1349

Qy 1321 GAATTTGGTCCAACTATAAGAAAAGCTCATACCTATATTAAGAATTCACAGGTTTTAGAA 1380
Db 1350 GAATATGGGCCCGTTTTGGAAAAAGCACATTCATTTGTCAAGAATTCCCAGGTGTTAGAA 1409

Qy 1381 GATTGTCCAGGTGATCTTAATAAATGGTACCGTCACATTTCAAAAGGTGCTTGGCCTTTT 1440
Db 1410 GACTGCCCTGGAGATCTGAATTACTGGTATCGCCACATTTCTAAAGGGGCTTGGCCTTTC 1469

Qy 1441 TCAACTGGAGATCATGGATGGCCAATTTCTGACTGCACAGCTGAAGGACTGAAAGCTGTT 1500
Db 1470 TCAACTGCAGATCACGGTTGGCCCATCTCTGACTGCACCGCAGAAGGACTGAAAGCTGCT 1529

Qy 1501 CTATTACTATCCAAAATTGCACCAGAAATACTTGGTGAGCCAATAGACGTGAAGCGATTA 1560
Db 1530 CTTTTGCTATCCAAAGTTCCCAAGGAGATTGTTGGTGAACCAATAGATGCAAAACGGTTA 1589

Qy 1561 TATGATTCTGTAAATGTCATTCTCTCACTACAGAATGAAGATGGTGGTTTTGCAACATAT 1620
Db 1590 TATGAAGCTGTTAATGTTATCATTTCTTTACAGAATGCAGATGGAGGCCTCGCAACATAT 1649

Qy 1621 GAGCTTAAACGATCTTATAATTGGTTGGAGATAATCAATCCTGCTGAAACTTTTGGTGAC 1680
Db 1650 GAGCTCACCAGGTCATACCCTTGGTTAGAGCTAATCAACCCAGCAGAAACCTTTGGCGAT 1709

Qy 1681 ATCGTTATTGATTATCCTTATGTGGAATGTACATCAGCAGCGATTCAAGCTTTGGCATCA 1740
Db 1710 ATTGTTATTGATTATCCTTACGTGGAATGTACATCAGCTGCTATCCAAGCTTTGATATCA 1769

Qy 1741 TTTAGGAAATTATATCCTGGGCATCGCCGACAAGAAATACAACATTGTATCGATAAAGCC 1800
Db 1770 TTTCGAAAGCTGTATCCTGGTCATCGAAAGAAGGAAGTAGATGAGTGCATTGAGAAGGCG 1829

Qy 1801 ACTACCTTCATTGAAAAAATACAAGCTTCAGATGGATCATGGTATGGTTCTTGGGGAGTT 1860
Db 1830 GTTAAGTTCATTGAATCCATTCAAGCAGCAGATGGCTCATGGTATGGATCATGGGCTGTT 1889

Qy 1861 TGCTTCACTTACGGTGCTTGGTTTGGGGTAAAAGGTCTGATTGCTGCTGGAAGGAGTTTC 1920
Db 1890 TGCTTCACGTATGGTACGTGGTTTGGAGTGAAAGGGCTGGTAGCTGTTGGAAAAACATTG 1949
```

```
Qy       982 ATTCTTTGGGCATCTCTACACAAGTTCCTTGAGCCTATTCTGATGCATTGGCCTGGAAAA 1041
Db       989 ATACTTTGGGCATCAATTGATAAGGTTTTGGAACCTATTTTAATGCGTTGGCCTGGAAAA 1048

Qy      1042 AGATTGAGGGAAAAGGCTATTATTTCTGCATTGGAGCATATACATTACGAAGATGAGAAT 1101
Db      1049 AAGTTGAGAGAGAAAGCTCTTCGCACCGTGATGGAACACATACATTATGAAGATGAAAAT 1108

Qy      1102 ACTCGATATATTTGCATAGGTCCTGTAAATAAGGTGTTAAATATGCTTTGCTGTTGGGTG 1161
Db      1109 ACTCGCTATATATGCATAGGGCCTGTAAACAAGGTGTTAAATATGCTATGCTGTTGGGCT 1168

Qy      1162 GAAGATCCAAATTCTGAGGCCTTCAAGTTGCATCTTCCCAGGATTTATGATTATCTATGG 1221
Db      1169 GAGGATCCAAACTCTGAGGCTTTCAAGCTGCATCTTCCAAGACTCTATGATTTTTTATGG 1228

Qy      1222 ATTGCAGAAGATGGCATGAAAATGCAGGGCTACAATGGAAGTCAACTATGGGACACTGCT 1281
Db      1229 CTTGCTGAGGATGGCATGAAAATGCAGGGGTACAATGGAAGTCAATTATGGGACACTGCT 1288

Qy      1282 TTTGCTGTCCAAGCAATTATTGCATCTAACCTCATTGAAGAATTTGGTCCAACTATAAGA 1341
Db      1289 TTTGCTGTTCAAGCAATCATATCAACTAGTTTGACTGAATATTGTGGCCCAACTCTCAGA 1348

Qy      1342 AAAGCTCATACCTATATTAAGAATTCACAGGTTTTAGAAGATTGTCCAGGTGATCTTAAT 1401
Db      1349 AAAGCACATACATTCATGAAAGACTCACAGGTGCTAGATGATTGTCCTGGTGATCTTGAT 1408

Qy      1402 AAATGGTACCGTCACATTTCAAAAGGTGCTTGGCCTTTTTCAACTGGAGATCATGGATGG 1461
Db      1409 TTTTGGTATCGTCATATATCAAAAGGTGCATGGCCTTTCTCCACTGCTGATCATGGATGG 1468

Qy      1462 CCAATTTCTGACTGCACAGCTGAAGGACTGAAAGCTGTTCTATTACTATCCAAAATTGCA 1521
Db      1469 CCCATTTCAGATTGTACAGCCGAAGGATTTAAGGCTGTGCTTCTGTTATCAAAATTCCCA 1528

Qy      1522 CCAGAAATAGTTGGTGAGCCAATAGACGTGAAGCGATTATATGATTCTGTAAATGTCATT 1581
Db      1529 GCAGAACTTGTTGGTGAACCATTAGATGCCAAGCGACTGTATGATGCTGTCAACGTTATT 1588

Qy      1582 CTCTCACTACAGAATGAAGATGGTGGTTTTGCAACATATGAGCTTAAACGATCTTATAAT 1641
Db      1589 CTTTCTTTACAGAATAGTGATGGTGGCTATGCAACATATGAACTCACAAGATCCTATCGC 1648

Qy      1642 TGGTTGGAGATAATCAATCCTGCTGAAACTTTTGGTGACATCGTTATTGATTATCCTTAT 1701
Db      1649 TGGCTAGAGTTAATCAACCCTGCCGAGACGTTCGGGGACATTGTTATTGACTATCCTTAT 1708

Qy      1702 GTGGAATGTACATCAGCAGCGATTCAAGCTTTGGCATCATTTAGGAAATTATATCCTGGG 1761
Db      1709 GCAGAATGTACATCAGCTGCAATTCAAGCGTTATCAGCTTTCAAAAAATTATATCCTGGG 1768

Qy      1762 CATCGCCGAGAAGAAATACAACATTGTATCGATAAAGCCACTACCTTCATTGAAAAAATA 1821
Db      1769 CACCGTCGGGAAGAGATACAGTGTTCTATTGAAAAAGCTGCAGACTTCATAGAGAAGATA 1828

Qy      1822 CAAGCTTCAGATGGATCATGGTATGGTTCTTGGGGAGTTTGCTTCACTTACGGTGCTTGG 1881
Db      1829 CAAGCTTCAGATGGCTCATGGTATGGCTCTTGGGGAGTTTGCTTCACCTATGGTACATGG 1888

Qy      1882 TTTGGGGTAAAAGGTCTGATTGCTGCTGGAAGGAGTTTCAGTAATTGCTCAAGCATCCGT 1941
Db      1889 TTTGGTGTAAAAGGGTTGATATCTGCTGGGAGAACCTTCTCTAATTGCTCAAGCATTCGC 1948

Qy      1942 AAAGCTTGTGAATTTCTGCTGTCCAAGCAGCTTCCTTCTGGTGGCTGGGGAGAGAGTTAT 2001
Db      1949 AAGGCTTGTTATTTTCTACTGTCCAAACAACTTGCTTCTGGTGGTTGGGGAGAGAGTTAC 2008

Qy      2002 CTGTCCTGTCAAAACAAGGTGTATTCAAATCTGGAAGGCAACAGGTCTCATGTGGTCAAC 2061
Db      2009 CTCTCCTGTCAGAACAAGGTGTATACAAATCTTGAGGGAGAACGTTCTCATGTTGTAAAT 2068

Qy      2062 ACTGGGTGGGCTATGTTGGCTCTCATTGATGCTGGACAGGCTAAGAGAGATTCGCAACCA 2121
Db      2069 ACTGGATGGGCTATGCTGGCTCTCATTGAGGCTGGGCAGGCAGAGAGAGATGCAACACCA 2128

Qy      2122 CTGCACCGGGCAGCTGCATACTTGATAAATTCCCAATTGGAGGACGGTGACTTTCCGCAG 2181
Db      2129 TTGCACCATGCAGCTAAATTATTGATTAATTCTCAAATGGAAAATGGAGATTTCCCTCAG 2188

Qy      2182 CAGGAAATAATGGGAGTCTTCAACAAGAATTGCATGATCACATACGCCGCATACAGAAAC 2241
Db      2189 GAGGAGATCATGGGGGTCTTCAACAAGAATTGTATGATCACTTATGCTGCGTACAGAAAC 2248

Qy      2242 ATATTCCCCATTTGGGCGTTGGGAGAATACCAATCCCAAGTATTGC 2287
Db      2249 ATTTTCCCTATTTGGGCTTTGGGGGACTACCGGTGTCGTGTACTTC 2294
```

RESULT 7
U02555

```
LOCUS       U02555                2561 bp    mRNA    linear   PLN 05-MAR-1998
DEFINITION  Arabidopsis thaliana Landsberg erecta cycloartenol synthase (CAS1)
            mRNA, complete cds.
ACCESSION   U02555 S67399
VERSION     U02555.1  GI:452445
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 2561)
  AUTHORS   Minet,M., Dufour,M.E. and Lacroute,F.
  TITLE     Complementation of Saccharomyces cerevisiae auxotrophic mutants by
            Arabidopsis thaliana cDNAs
  JOURNAL   Plant J. 2 (3), 417-422 (1992)
  MEDLINE   93272056
   PUBMED   1303803
REFERENCE   2  (bases 1 to 2561)
  AUTHORS   Corey,E.J., Matsuda,S.P. and Bartel,B.
  TITLE     Isolation of an Arabidopsis thaliana gene encoding cycloartenol
            synthase by functional expression in a yeast mutant lacking
            lanosterol synthase by the use of a chromatographic screen
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 90 (24), 11628-11632 (1993)
  MEDLINE   94089686
   PUBMED   7505443
REFERENCE   3  (bases 1 to 2561)
  AUTHORS   Matsuda,S.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-OCT-1993) Department of Chemistry, Harvard
            University, Cambridge, MA 02138, USA
COMMENT     On Mar 5, 1998 this sequence version replaced gi:455864.
FEATURES             Location/Qualifiers
     source          1. .2561
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /cultivar="Landsberg erecta ecotype"
                     /db_xref="taxon:3702"
                     /clone_lib="cDNA from Minet et al"
                     /dev_stage="seedling"
     gene            1. .2561
                     /gene="CAS1"
     CDS             54. .2333
                     /gene="CAS1"
                     /EC_number="5.4.99.8"
                     /function="cycloartenol synthesis; 2,3-oxidosqualene
                     metabolism"
                     /note="(S)-2,3-epoxysqualene mutase"
                     /codon_start=1
                     /product="cycloartenol synthase"
                     /protein_id="AAC04931.1"
                     /db_xref="GI:452446"
                     /translation="MWKLKIAEGGSPWLRTTNNHVGRQFWEFDPNLGTPEDLAAVEEA
                     RKSFSDNRFVQKHSADLLMRLQFSRENLISPVLPQVKIEDTDDVTEEMVETTLKRGLD
```

```
Qy       2004 GTCCTGTCAAAACAAGGTGTATTCAAATCTGGAAGGCAACAGGTCTCATGTGGTCAACAC 2063
Db       2064 TTCCTGTCAAAACAAGGTGTACACAAATCTTGAGGGTAACCGTTCTCATGTTGTCAATAC 2123

Qy       2064 TGGGTGGGCTATGTTGGCTCTCATTGATGCTGGACAGGCTAAGAGAGATTCGCAACCACT 2123
Db       2124 TGGATGGGCTATGCTGGCTCTCATCGATGCTGGACAGGCCGAGAGAGACGCAACACCATT 2183

Qy       2124 GCACCGGGCAGCTGCATACTTGATAAATTCCCAATTGGAGGACGGTGACTTTCCGCAGCA 2183
Db       2184 GCACCGTGCAGCCAAATTATTAATTAATTCTCAAATGGAAAATGGTGATTTCCCACAGGA 2243

Qy       2184 GGAAATAATGGGAGTCTTCAACAAGAATTGCATGATCACATACGCCGCATACAGAAACAT 2243
Db       2244 GGAGATCATGGGGGTCTTCGACAAGAATTGCATGATCACTTATGCTGCATACAGAAACAT 2303

Qy       2244 ATTCCCCATTTGGGCGTTGGGAGAATACCAATCCCAAGTATTGCAATCTCGTTAATCGAG 2303
Db       2304 TTTCCCTATTTGGGCTTTGGGAGAATATCGATGTCGAGTACTCCAGGGGCCATCATGAAA 2363

Qy       2304 CCTTAGTT 2311
Db       2364 ATTTATTT 2371
```

RESULT 6
AY520819
LOCUS AY520819 2547 bp mRNA linear PLN 03-FEB-2004
DEFINITION Centella asiatica cycloartenol synthase (OSCCCS) mRNA, complete cds.
ACCESSION AY520819
VERSION AY520819.1 GI:41387167
KEYWORDS .
SOURCE Centella asiatica
ORGANISM Centella asiatica
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; campanulids; Apiales; Mackinlayaceae; Centella.
REFERENCE 1 (bases 1 to 2547)
AUTHORS Kim,O.T., Kim,M.Y., Ahn,J.C. and Hwang,B.
TITLE Molecular cloning and characterization of cDNAs encoding oxidosqualene cyclase from Centella asiatica
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2547)
AUTHORS Kim,O.T., Kim,M.Y., Ahn,J.C. and Hwang,B.
TITLE Direct Submission
JOURNAL Submitted (08-JAN-2004) Department of Biology, Chonnam National University, Gwangju 500-757, Korea
FEATURES Location/Qualifiers
source 1. .2547
/organism="Centella asiatica"
/mol_type="mRNA"
/db_xref="taxon:48106"
gene 1. .2547
/gene="OSCCCS"
CDS 32. .2305
/gene="OSCCCS"
/note="oxidosqualene cyclase"
/codon_start=1
/product="cycloartenol synthase"
/protein_id="AAS01524.1"
/db_xref="GI:41387168"
/translation="MWKLKVAEGGNPWLRTVNNHVGRQIWEFDPKLGSPEELAEIEKA
RENFHNHRFEKQHSSDLLMRLQFANENPRHEVLPQVKVKDIGDISEDKVTITLKRALS
FYSTLQANDGHWAGDYGGPMFLMPGLVITLSITEALNAILSKEHKREICRYLYNHQNR
DGGWGLHIEGPSTMFGSVLNYVTLRLLGEGTKDGQGAMEKGRQWILDHGGATAITSWG
RMWLSVLGVFEWSGNNPLPPEIWLFPYNLPFHPGRMWCHCRMVYLPMSYLYGKRFVGP
ITPTVLSLRKELFTVPYHEIDWNQARNLCAKEDLYYPHPLVQDILWASIDKVLEPILM
RWPGKKLREKALRTVMEHIHYEDENTRYICIGPVNKVLNMLCCWAEDPNSEAFKLHLP
RLYDFLWLAEDGMKMQGYNGSQLWDTAFAVQAIISTSLTEYCGPTLRKAHTFMKDSQV
LDDCPGDLDFWYRHISKGAWPFSTADHGWPISDCTAEGFKAVLLLSKFPAELVGEPLD
AKRLYDAVNVILSLQNSDGGYATYELTRSYRWLELINPAETFGDIVIDYPYAECTSAA
IQALSAFKKLYPGHRREEIQCSIEKAADFIEKIQASDGSWYGSWGVCFTYGTWFGVKG
LISAGRTFSNCSSIRKACYFLLSKQLASGGWGESYLSCQNKVYTNLEGERSHVVNTGW
AMLALIEAGQAERDATPLHHAAKLLINSQMENGDFPQEEIMGVFNKNCMITYAAYRNI
FPIWALGDYRCRVLLAP"
ORIGIN

```
Query Match 58.6%; Score 1501.2; DB 8; Length 2547;
Best Local Similarity 78.9%; Pred. No. 0;
Matches 1788; Conservative 0; Mismatches 478; Indels 0; Gaps 0;

Qy         22 AGAATGTGGAAGCTCAAGTTCGCCGAAGGAGGGAATCCATGGCTTCGGACATTGAACAAT 81
Db         29 AAAATGTGGAAGCTCAAAGTCGCTGAAGGAGGAAATCCATGGCTTCGGACTGTCAACAAT 88

Qy         82 CACGTTGGAAGACAGGTGTGGGAGTTCGATCCTAAGCTTGGATCGCCGCAAGATCTCCTC 141
Db         89 CACGTCGGCCGGCAAATTTGGGAGTTCGATCCAAAACTTGGATCTCCCGAAGAGCTTGCG 148

Qy        142 GAGATTGAGAAAGCTCGCCAGAATTTTCACGATAACCGCTTTACCCACAAACACAGCGCT 201
Db        149 GAGATCGAGAAAGCTCGCGAGAATTTTCACAATCATCGATTCGAGAAACAGCACAGTTCC 208

Qy        202 GATCTACTTATGCGGATGCAGTTCGCAAGAGAGAACCCAACACGTGAAGTCTTGCCCAAA 261
Db        209 GATCTACTGATGCGGCTACAGTTCGCCAATGAGAATCCACGCCATGAAGTCTTACCCCAA 268

Qy        262 GTCGGAGTTAAGGATATTGAGGATGTGACCCAAGAGATTGTGACAAAAACATTAAGAAGG 321
Db        269 GTTAAAGTAAAAGATATCGGAGATATTTCGGAGGATAAAGTAACAATTACGTTAAAAAGA 328

Qy        322 GCCGTAAGTTTCCATTCAACTCTCCAGTGCCATGACGGACACTGGCCGGGAGATTATGGA 381
Db        329 GCTTTGAGTTTTTATTCAACTCTTCAAGCCAATGATGGACATTGGGCAGGAGATTATGGT 388

Qy        382 GGTCCCATGTTTCTGATGCCTGGCTTGGTAATTACTCTGTCTATCACTGGGGCGTTGAAT 441
Db        389 GGTCCTATGTTTTTAATGCCTGGTTTGGTTATCACTCTATCAATTACTGAGGCACTGAAT 448

Qy        442 ACAGTCTTAACTGAAGAACATAGAAAGGAAATATGCCGTTACCTCTATAATCATCAAAAC 501
Db        449 GCTATCTTGTCCAAAGAACACAAACGTGAGATATGCCGCTATCTTTACAATCATCAGAAT 508

Qy        502 AAGGATGGTGGGTGGGGTTTGCATATTGAAGGTCCAAGCACCATGTTTGGCTCTGTCTTG 561
Db        509 AGAGATGGTGGGTGGGGTTTACACATTGAAGGTCCAAGCACTATGTTTGGTAGTGTCTTG 568

Qy        562 AGTTATATTACTCTGAGATTGCTAGGTGAGGGGCCTAATGATGGACAAGGGGAAATGGAG 621
Db        569 AACTATGTTACTCTGAGATTGCTTGGTGAAGGAACTAAGGATGGGCAAGGAGCAATGGAG 628

Qy        622 AAGGCACGTGACTGGATTCTAGGGCATGGTGGTGCTACTTATATAACGTCATGGGGGAAG 681
Db        629 AAAGGGCGTCAGTGGATTCTGGATCATGGTGGTGCCACTGCAATTACATCATGGGGAAGA 688

Qy        682 ATGTGGCTTTCAGTACTTGGAGTGTATGAATGGTCTGGAAATAATCCCCTGCCCCCTGAG 741
Db        689 ATGTGGCTTTCAGTTCTTGGTGTATTTGAATGGTCTGGAAATAATCCCCTCCCCCCTGAG 748

Qy        742 ATATGGCTCCTTCCATACATGCTTCCATTTCATCCAGGAAGGATGTGGTGTCACTGCCGG 801
Db        749 ATATGGTTATTTCCATATAATCTTCCATTCCACCCAGGAAGGATGTGGTGCCACTGTCGG 808

Qy        802 ATGGTCTATTTGCCGATGTCCTACTTATATGGCAAGAGGTTTGTTGGTCCAATCTCACCA 861
Db        809 ATGGTTTATCTTCCTATGTCATACTTATATGGGAAGAGGTTTGTTGGTCCAATCACACCG 868

Qy        862 ACAGTATTATCTTTGAGAAAAGAGCTTTATACAGTACCATACCATGATATAGATTGGGAT 921
Db        869 ACTGTCCTTTCTTTGAGAAAGGAGCTCTTCACTGTCCCTTATCATGAAATAGATTGGAAT 928

Qy        922 CAGGCTCGCAATTTGTGTGCAAAGGAAGATTTGTACTATCCTCACCCACTTGTACAGGAT 981
Db        929 CAAGCACGCAACCTATGTGCAAAGGAAGATCTGTACTATCCCACACCCTCTTGTGCAGGAT 988
```

ITPTVLSLRKEVFSVPYHEIDWNQARNLCAKEDLYYPHPLIQDILWASLDKVWEPIFM
HWPAKKLREKSLRTVMEHIHYEDENTRYICIGPVNKVLNMLCCWVEDPNSEAFKLHLP
RLHDFLWLAEDGMKMQGYNGSQLWDTAFAVQAIISTNLAEEYGPTLRKAHTFMKNSQV
LDDCPGDLDAWYRHVSKGAWPFSTADHGWPISDCTAEGFKAVLQLSKLPSELVGEPLD
AKRLYDAVNVILSLQNSDGGYATYELTRSYSWLELVNPAETFGDIVIDYPYVECTSAA
IQALTAFKKLFPGHRREEIQHSIEKAALFIEKIQSSDGSWYGSWGVCFTYGTWFGIKG
LVTAGRTFSSCASIRKACDFLLSKQVASGGWGESYLSCQNKVYTNLEGNRSHVVNTGW
AMLALIDAGQAERDATPLHRAAKLLINSQMENGDFPQEEIMGVFDKNCMITYAAYRNI
FPIWALGEYRCRVLQGPS"

ORIGIN

Query Match 58.7%; Score 1502.4; DB 8; Length 2559;
Best Local Similarity 78.5%; Pred. No. 0;
Matches 1797; Conservative 0; Mismatches 491; Indels 0; Gaps 0;

```
Qy        24 AATGTGGAAGCTCAAGTTCGCCGAAGGAGGGAATCCATGGCTTCGGACATTGAACAATCA 83
             ||||||||||||  ||  |  ||  |||||  ||||||||||||||  ||  |||||  ||||
Db        84 AATGTGGAAGCTTAAAATTGCAGAGGGAGGAAATCCATGGCTTCGAACTCTGAACGATCA 143

Qy        84 CGTTGGAAGACAGGTGTGGGAGTTCGATCCTAAGCTTGGATCGCCGCAAGATCTCCTCGA 143
             |||  ||   |  ||   | ||||||||||||||  ||   ||||||||||   ||||  ||     ||
Db       144 CGTCGGCCGGCAAATTTGGGAGTTCGATCCCAACATTGGATCGCCTGAAGAGCTTGCGGA 203

Qy       144 GATTGAGAAAGCTCGCCAGAATTTTCACGATAACCGCTTTACCCACAAACACAGCGCTGA 203
             | | |||||||  |||   | |||||||  |  ||  |  ||  ||      | ||  |||||  ||  ||
Db       204 GGTCGAGAAAGTTCGTGAAAATTTTCGCAATCATCGATTCGAGAAAAAGCACAGTGCCGA 263

Qy       204 TCTACTTATGCGGATGCAGTTCGCAAGAGAGAACCCAACACGTGAAGTCTTGCCCAAAGT 263
             |||  ||  |||||  ||  |||||  ||||   |||||  |||    |||  |||||  |||  ||||
Db       264 TCTCCTCATGCGCATTCAGTTTGCAAATGAGAATCCAGGCAGTGTAGTCTTACCCCAAGT 323

Qy       264 CGGAGTTAAGGATATTGAGGATGTGACCCAAGAGATTGTGACAAAAACATTAAGAAGGGC 323
             |   |||  ||  |||   |||  |||  |   |  ||  |   ||  |||    |||||||  |||  ||
Db       324 CAAAGTAAATGATGGTGAAGATATTTCGGAGGATAAAGTAACAGTTACATTAAAAAGAGC 383

Qy       324 CGTAAGTTTCCATTCAACTCTCCAGTGCCATGACGGACACTGGCCGGGAGATTATGGAGG 383
               |  ||||||  ||||||||||  |||    |||||  |||||  |||||  ||||||||||  ||
Db       384 TATGAGTTTCTATTCAACTCTTCAGGCACATGATGGACATTGGCCAGGAGATTATGGTGG 443

Qy       384 TCCCATGTTTCTGATGCCTGGCTTGGTAATTACTCTGTCTATCACTGGGGCGTTGAATAC 443
             |||  ||||||||  |||||||  |||||  ||  ||  |||||||  |  |||||
Db       444 TCCTATGTTTCTAATGCCTGGTTTGGTTATTACGCTATCGATAACTGGGGTGCTGAATGT 503

Qy       444 AGTCTTAACTGAAGAACATAGAAAGGAAATATGCCGTTACCTCTATAATCATCAAAACAA 503
             |||  ||   |  |||||||||  |    ||  ||  ||||||||  ||  ||  |||||||  ||||
Db       504 AGTTTTGTCCAAAGAACATAAACGTGAGATCTGCCGTTATCTTTACAATCATCAGAACAG 563

Qy       504 GGATGGTGGGTGGGGTTTGCATATTGAAGGTCCAAGCACCATGTTTGGCTCTGTCTTGAG 563
               ||||||||||||||||  ||  ||||||||||||||  ||  ||||||||  |||||||||
Db       564 AGATGGTGGGTGGGGTTTACACATTGAAGGTCCAAGTACTATGTTTGGTACTGTCTTGAA 623

Qy       564 TTATATTACTCTGAGATTGCTAGGTGAGGGGCCTAATGATGGACAAGGGGAAATGGAGAA 623
             |||  |||||  ||||  |||||  ||||  ||  ||||||||||  |||||||  ||||||  ||
Db       624 CTATGTTACTTTGAGGTTGCTTGGTGAAGGTGCTAATGATGGCCAAGGGGCAATGGAAAA 683

Qy       624 GGCACGTGACTGGATTCTAGGGCATGGTGGTGCTACTTATATAACGTCATGGGGGAAGAT 683
               |  |||  |  ||||||||  |  ||||||  ||||  |||    |||||  ||||||||||  ||
Db       684 AGGGCGTCAATGGATTCTGGATCATGGTAGTGCCACTGCAATAACATCATGGGGGAAAAT 743

Qy       684 GTGGCTTTCAGTACTTGGAGTGTATGAATGGTCTGGAAATAATCCCCTGCCCCCTGAGAT 743
             ||||||||||||  |||||  ||  |||||||||||||||||||||||||||||||||||||
Db       744 GTGGCTTTCAGTTCTTGGTGTATTTGAATGGTCTGGAAATAATCCCCTGCCCCCTGAGAC 803

Qy       744 ATGGCTCCTTCCATACATGCTTCCATTTCATCCAGGAAGGATGTGGTGTCACTGCCGGAT 803
             |||||  ||||||||  ||  ||||||  |  ||  ||||||||||||||||  |||  |  |||||
Db       804 GTGGCTGCTTCCATATATTCTTCCAATCCACCCAGGAAGGATGTGGTGCCACCGTCGGAT 863

Qy       804 GGTCTATTTGCCGATGTCCTACTTATATGGCAAGAGGTTTGTTGGTCCAATCTCACCAAC 863
             |||  |||  ||||  |||||  ||||||||||  |||||  ||||||||||||||||  ||||  ||
Db       864 GGTTTATCTGCCTATGTCATACTTATATGGGAAGAGATTTGTTGGTCCAATCACACCGAC 923

Qy       864 AGTATTATCTTTGAGAAAAGAGCTTTATACAGTACCATACCATGATATAGATTGGGATCA 923
               |||  |  ||||||||||||  |||  |  |  |  ||  ||  ||  |||||  |||||||||  ||||

Db       924 TGTACTTTCTTTGAGAAAGGAGGTCTTCAGTGTCCCTTATCATGAAATAGATTGGAATCA 983

Qy       924 GGCTCGCAATTTGTGTGCAAAGGAAGATTTGTACTATCCTCACCCACTTGTACAGGATAT 983
             |||||  ||  |  |||||  |||||||||  ||||||||||  |||||  ||  ||||||||||
Db       984 AGCTCGTAACCTTTGTGCGAAGGAAGATCTGTACTATCCACACCCTCTCATACAGGATAT 1043

Qy       984 TCTTTGGGCATCTCTACACAAGTTCCTTGAGCCTATTCTGATGCATTGGCCTGGAAAAAG 1043
             ||||||||||||  ||  |||||  ||  ||  ||||||  |  ||||||||||||||  ||  |
Db      1044 ACTTTGGGCATCACTGGACAAGGTCTGGGAACCTATTTTCATGCATTGGCCTGCGAAGAA 1103

Qy      1044 ATTGAGGGAAAAGGCTATTATTTCTGCATTGGAGCATATACATTACGAAGATGAGAATAC 1103
             |||||  ||  |||  ||  ||     |||    ||||  ||  ||||||||  |||||  ||||||||
Db      1104 GTTGAGAGAGAAGTCTCTTCGCACTGTGATGGAACACATACATTATGAAGACGAGAATAC 1163

Qy      1104 TCGATATATTTGCATAGGTCCTGTAAATAAGGTGTTAAATATGCTTTGCTGTTGGGTGGA 1163
             |||  |||||  ||  |||||  ||||||||  ||||||||||||||||||  ||||||||||  ||
Db      1164 TCGTTATATATGTATAGGGCCTGTAAACAAGGTGTTAAATATGCTATGCTGTTGGGTTGA 1223

Qy      1164 AGATCCAAATTCTGAGGCCTTCAAGTTGCATCTTCCCAGGATTTATGATTATCTATGGAT 1223
             ||||||  ||  ||  |||||  ||||||||  |||||  ||  ||  |  ||||||  |  |  |||  |
Db      1224 AGATCCGAACTCCGAGGCTTTCAAGTTACATCTCCCAAGACTCCATGATTTTTTGTGGCT 1283

Qy      1224 TGCAGAAGATGGCATGAAAATGCAGGGCTACAATGGAAGTCAACTATGGGACACTGCTTT 1283
             |||  ||  |||||  ||||||||||||||  ||  ||||||||||||  |  ||||||||||||||
Db      1284 TGCTGAGGATGGTATGAAAATGCAGGGGTATAATGGAAGTCAATTGTGGGACACTGCTTT 1343

Qy      1284 TGCTGTCCAAGCAATTATTGCATCTAACCTCATTGAAGAATTTGGTCCAACTATAAGAAA 1343
             |||  ||  ||||||||||||  |  |||||  ||||||||  |||||||||  |  |||||
Db      1344 TGCCGTTCAAGCAATTATTTCGACAAACCTTGCTGAAGAATACGGTCCAACTCTTAGAAA 1403

Qy      1344 AGCTCATACCTATATTAAGAATTCACAGGTTTTAGAAGATTGTCCAGGTGATCTTAATAA 1403
             |||  |||||  |  ||  ||  ||  |||||  ||||  |||||  ||  ||  ||||||  ||
Db      1404 AGCACATACGTTCATGAAAAACTCGCAGGTGCTAGATGATTGCCCCGGGGATCTTGATGC 1463

Qy      1404 ATGGTACCGTCACATTTCAAAAGGTGCTTGGCCTTTTTCAACTGGAGATCATGGATGGCC 1463
             |||||  ||  ||  ||||||||||||||  |||||  ||  |||||||  ||||||  ||||||||
Db      1464 CTGGTATCGCCATGTTTCAAAAGGTGCATGGCCGTTCTCAACTGCAGATCACGGATGGCC 1523

Qy      1464 AATTTCTGACTGCACAGCTGAAGGACTGAAAGCTGTTCTATTACTATCCAAAATTGCACC 1523
             |||||  |||||  ||  ||  ||||||  |  ||  |||||  ||  |  ||||  |||  |  ||  |
Db      1524 CATTTCAGACTGTACTGCCGAAGGATTCAAGGCTGTGCTTCAATTATCAAAATTGCCATC 1583

Qy      1524 AGAAATAGTTGGTGAGCCAATAGACGTGAAGCGATTATATGATTCTGTAAATGTCATTCT 1583
             ||||  |  ||||||||  |||  ||||  |  |||||  |  ||||||  ||||  ||  ||  |||||
Db      1584 AGAACTTGTTGGTGAACCATTAGATGCCAAGCGGCTGTATGATGCTGTCAACGTTATTCT 1643

Qy      1584 CTCACTACAGAATGAAGATGGTGGTTTTGCAACATATGAGCTTAAACGATCTTATAATTG 1643
             ||  ||||||||  ||||||||  |  ||||||||||||  ||  |  ||||  ||||  ||
Db      1644 TTCTTTACAGAATAGTGATGGTGGCTATGCAACATATGAACTAACGAGATCCTATAGCTG 1703

Qy      1644 GTTGGAGATAATCAATCCTGCTGAAACTTTTGGTGACATCGTTATTGATTATCCTTATGT 1703
             |||||||  ||  ||||  ||||||||  ||  ||||||||  ||||||||  ||  ||||||||
Db      1704 GTTGGAGTTAGTCAACCCTGCTGAGACGTTCGGTGACATTGTTATTGACTACCCTTATGT 1763

Qy      1704 GGAATGTACATCAGCAGCGATTCAAGCTTTGGCATCATTTAGGAAATTATATCCTGGGCA 1763
             ||||||||||||||  ||  ||||||||||||  |  ||  |  ||||||||  ||||||||
Db      1764 AGAATGTACATCAGCTGCAATTCAAGCTTTGACAGCTTTCAAGAAATTATTTCCTGGGCA 1823

Qy      1764 TCGCCGAGAAGAAATACAACATTGTATCGATAAAGCCACTACCTTCATTGAAAAAATACA 1823
             |||  ||||||||||||||  ||||  |||  ||  |||||  ||  ||||||||||  |||||
Db      1824 TCGTCGAGAAGAAATACAGCATTCTATTGAAAAAGCTGCTCTGTTCATTGAAAAGATACA 1883

Qy      1824 AGCTTCAGATGGATCATGGTATGGTTCTTGGGGAGTTTGCTTCACTTACGGTGCTTGGTT 1883
             | |  ||||||||  ||||||||  ||  ||||||||||  |||||||  ||  ||  |||||
Db      1884 ATCATCAGATGGCTCATGGTACGGCTCTTGGGGAGTCTGCTTCACCTATGGCACATGGTT 1943

Qy      1884 TGGGGTAAAAGGTCTGATTGCTGCTGGAAGGAGTTTCAGTAATTGCTCAAGCATCCGTAA 1943
             |||  ||||||||  |  |||||||  ||||  |||  ||  ||||  |  |||||  ||  ||
Db      1944 TGGTATAAAAGGTTTAGTAACTGCTGGGAGGACCTTCTCTAGTTGCGCTAGCATTCGCAA 2003

Qy      1944 AGCTTGTGAATTTCTGCTGTCCAAGCAGCTTCCTTCTGGTGGCTGGGGAGAGAGTTATCT 2003
             ||||||||  |||||  ||||||||  ||  ||  ||||||||||  ||||||||||||||||  ||
Db      2004 GGCTTGTGATTTTCTATTGTCCAAACAAGTTGCTTCTGGTGGTTGGGGAGAGAGTTACCT 2063
```

```
Db        982 TTGGTCCAATCACGCCAACAGTTATGTCTTTGAGGAAGGAGCTTTATTCTGTCCCCTATC 1041
Qy        905 ATGATATAGATTGGGATCAGGCTCGCAATTTGTGTGCAAAGGAAGATTTGTACTATCCTC 964
Db       1042 ATGAAATAGATTGGAATCAGGCACGCAACCTATGTGCAAAGGAAKATCTGTACTACCCAC 1101
Qy        965 ACCCACTTGTACAGGATATTCTTTGGGCATCTCTACACAAGTTCCTTGAGCCTATTCTGA 1024
Db       1102 ATCCGCTGGTGCAGGATATACTTTGGGCATCTCTTCATAAGCTTGTTGAACCGGTTTTAA 1161
Qy       1025 TGCATTGGCCTGGAAAAAGATTGAGGGAAAAGGCTATTATTTCTGCATTGGAGCATATAC 1084
Db       1162 TGCGTTGGCCTGGAAAAAGACTGAGAGAAAAGGCTCTTCGCACTGTGTTGGAACATATAC 1221
Qy       1085 ATTACGAAGATGAGAATACTCGATATATTTGCATAGGTCCTGTAAATAAGGTGTTAAATA 1144
Db       1222 ACTATGAAGATGAAAATACTCGTTATATTTGCATAGGACCTGTAAACAAGGTGTTAAATA 1281
Qy       1145 TGCTTTGCTGTTGGGTGGAAGATCCAAATTCTGAGGCCTTCAAGTTGCATCTTCCCAGGA 1204
Db       1282 TGCTCTGTTGTTGGGTGGAAGATCCAAATTCCGAGGCTTTTAAGTTGCATCTTCCAAGAA 1341
Qy       1205 TTTATGATTATCTATGGATTGCAGAAGATGGCATGAAAATGCAGGGCTACAATGGAAGTC 1264
Db       1342 TAAATGACTATTTGTGGATTGCTGAAGATGGAATGAAAATGCAGGGTTATAATGGAAGTC 1401
Qy       1265 AACTATGGGACACTGCTTTTGCTGTCCAAGCAATTATTGCATCTAACCTCATTGAAGAAT 1324
Db       1402 AATTATGGGATACTGCCTTTGCTGTTCAAGCTATTATTTCAACCAACCTTTTTGAAGAAT 1461
Qy       1325 TTGGTCCAACTATAAGAAAAGCTCATACCTATATTAAGAATTCACAGGTTTTAGAAGATT 1384
Db       1462 ATGGTCCAACTCTCGAAAAAGCACATATGTACATAAAAAAATCGCAGGTTCGAGAAGATT 1521
Qy       1385 GTCCAGGTGATCTTAATAAATGGTACCGTCACATTTCAAAAGGTGCTTGGCCTTTTTCAA 1444
Db       1522 GCCCAGGTGATCTTGATTTTTGGTATCGCCACATTTCAAAAGGCGCTTGGCCTTTCTCAA 1581
Qy       1445 CTGGAGATCATGGATGGCCAATTTCTGACTGCACAGCTGAAGGACTGAAAGCTGTTCTAT 1504
Db       1582 CTGCCGATCACGGATGGCCCATATCAGACTGCACAGCAGAAGGATTGAAAGCTGCTCTTT 1641
Qy       1505 TACTATCCAAAATTGCACCAGAAATAGTTGGTGAGCCAATAGACGTGAAGCGATTATATG 1564
Db       1642 TATTATCTAAAATTCCGCCAGATGTTGTTGGTGAACCATTAGTTGAAGAGCGATTGTATG 1701
Qy       1565 ATTCTGTAAATGTCATTCTCTCACTACAGAATGAAGATGGTGGTTTTGCAACATATGAGC 1624
Db       1702 ATGCTGTAAATGTCATTCTCTCATTACAGAATGCGGATGGTGGCTTTGCAACATATGAGC 1761
Qy       1625 TTAAACGATCTTATAATTGGTTGGAGATAATCAATCCTGCTGAAACTTTTGGTGACATCG 1684
Db       1762 TCACAAGATCCTATCCCTGGTTGGAGTTGATCAACCCTGCTGAAACCTTTGGGGACATTG 1821
Qy       1685 TTATTGATTATCCTTATGTGGAATGTACATCAGCAGCGATTCAAGCTTTGGCATCATTTA 1744
Db       1822 TTATTGATTATAATTATGTGGAATGTACGTCAGCAGCAATTCAAGCTCTGACATCGTTTA 1881
Qy       1745 GGAAATTATATCCTGGGCATCGCCGAGAAGAAATACAACATTGTATCGATAAAGCCACTA 1804
Db       1882 AGAAATCATATCCCAAGCATCGGGAGGAGGAGGTAGACGTCTGTATCAAAAGAGCTGCTA 1941
Qy       1805 CCTTCATTGAAAAAATACAAGCTTCAGATGGATCATGGTATGGTTCTTGGGGAGTTTGCT 1864
Db       1942 TGTTCACTGAAAAGATACAAGCATCAGATGGTTCATGGTATGGTTCTTGGGGTGTTTGCT 2001
Qy       1865 TCACTTACGGTGCTTGGTTTGGGGTAAAAGGTCTGATTGCTGCTGGAAGGAGTTTCAGTA 1924
Db       2002 TCACTTACGGTACATGGTTTGGGGTAAAAGGGTTAGTGGCTGCTGGAAAGAACTTCAATG 2061
Qy       1925 ATTGCTCAAGCATCCGTAAAGCTTGTGAATTTCTGCTGTCCAAGCAGCTTCCTTCTGGTG 1984
Db       2062 ATTGCTTTGGTATCCGCAAAGCTTGTGATTTTCTGCTGTCTAAGCAGCTTCCTTCAGGTG 2121
```

```
Qy       1985 GCTGGGGAGAGAGTTATCTGTCCTGTCAAAACAAGGTGTATTCAAATCTGGAAGGCAACA 2044
Db       2122 GCTGGGGAGAGAGTTATCTTTCATGTCAAAACAAGGTGTATTCACATGTAGAGGGCAACA 2181
Qy       2045 GGTCTCATGTGGTCAACACTGGGTGGGCTATGTTGGCTCTCATTGATGCTGGACAGGCTA 2104
Db       2182 GGTCTCATGTTGTAAATACTGGATGGGCTATGTTGGCGCTTATAGAAGCTGGACAGGCTG 2241
Qy       2105 AGAGAGATTCGCAACCACTGCACCGGGCAGCTGCATACTTGATAAATTCCCAATTGGAGG 2164
Db       2242 AGAGAGACCCAACGCCATTGCACCGTGCAGCTAGGGTTTTGATAAATTCTCAAATGGAGA 2301
Qy       2165 ACGGTGACTTTCCGCAGCAGGAAATAATGGGAGTCTTCAACAAGAATTGCATGATCACAT 2224
Db       2302 ATGGAGATTTTCCTCAAGAGGAAATCATGGGGGTCTTCAATAGGAACTGTATGATTACCT 2361
Qy       2225 ACGCCGCATACAGAAACATATTCCCCATTTGGGCGTTGGGAGAATACCAATCCCAAGTAT 2284
Db       2362 ACGCCGCTTACAGGAACATTTTTCCTATTTGGGCATTGGGAGAGTACCGCTGTCGGGTAT 2421
Qy       2285 TGCAATCTCGTTAATCGAGC 2304
Db       2422 TGCAGGCTCCCTAAGCTAAC 2441
```

RESULT 5
AB009029

```
LOCUS       AB009029               2559 bp    mRNA    linear   PLN 03-OCT-1998
DEFINITION  Panax ginseng OSCPNX1 mRNA for Cycloartenol Synthase, complete cds.
ACCESSION   AB009029
VERSION     AB009029.1  GI:3688597
KEYWORDS    OSCPNX1; Cycloartenol Synthase.
SOURCE      Panax ginseng
  ORGANISM  Panax ginseng
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; campanulids; Apiales; Araliaceae; Panax.
REFERENCE   1  (sites)
  AUTHORS   Kushiro,T., Shibuya,M. and Ebizuka,Y.
  TITLE     Beta-amyrin synthase--cloning of oxidosqualene cyclase that
            catalyzes the formation of the most popular triterpene among higher
            plants
  JOURNAL   Eur. J. Biochem. 256 (1), 238-244 (1998)
  MEDLINE   98417451
   PUBMED   9746369
REFERENCE   2  (bases 1 to 2559)
  AUTHORS   Kushiro,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-NOV-1997) Tetsuo Kushiro, The University of Tokyo,
            Graduate School of Pharmaceutical Sciences; 7-3-1 Hongo, Bunkyo-ku,
            Tokyo 113, Japan (E-mail:tkushiro@mol.f.u-tokyo.ac.jp,
            Tel:03-3812-2111, Fax:03-5689-2982)
COMMENT     Sequence updated (15-May-1998).
FEATURES             Location/Qualifiers
     source          1. .2559
                     /organism="Panax ginseng"
                     /mol_type="mRNA"
                     /db_xref="taxon:4054"
                     /tissue_type="Hairy Root"
     gene            1. .2559
                     /gene="OSCPNX1"
     CDS             85. .2361
                     /gene="OSCPNX1"
                     /codon_start=1
                     /product="Cycloartenol Synthase"
                     /protein_id="BAA33460.1"
                     /db_xref="GI:3688598"
                     /translation="MWKLKIAEGGNPWLRTLNDHVGRQIWEFDPNIGSPEELAEVEKV
                     RENFRNHRFEKKHSADLLMRIQFANENPGSVVLPQVKVNDGEDISEDKVTVTLKRAMS
                     FYSTLQAHDGHWPGDYGGPMFLMPGLVITLSITGVLNVVLSKEHKREICRYLYNHQNR
                     DGGWGLHIEGPSTMFGTVLNYVTLRLLGEGANDGQGAMEKGRQWILDHGSATAITSWG
                     KMWLSVLGVFEWSGNNPLPPETWLLPYILPIHPGRMWCHRRMVYLPMSYLYGKRFVGP
```

```
Db       2033 AAAGCTTGTGAGTTTCTGCTGTCTAAGCAGCTCCCATCTGGTGGTTGGGCAGAAAGTTAT 2092

Qy       2002 CTGTCCTGTCAAAACAAGGTGTATTCAAATCTGGAAGGCAACAGGTCTCATGTGGTCAAC 2061

Db       2093 CTGTCATGTCAAAACAAGGTGTATTCAAATTTGGAAGGCAACAGGTCCCATGTAGTAAAC 2152

Qy       2062 ACTGGGTGGGCTATGTTGGCTCTCATTGATGCTGGACAGGCTAAGAGAGATTCGCAACCA 2121

Db       2153 ACTGGGTGGGCTATGCTGGCTCTCATTGAGGCTGAACAGGCAAAGAGAGATCCAACACCA 2212

Qy       2122 CTGCACCGGGCAGCTGCATACTTGATAAATTCCCAATTGGAGGACGGTGACTTTCCGCAG 2181

Db       2213 CTGCACCGTGCAGCTGTATGCTTGATAAATTCCCAATTGGAGAATGGTGACTTTCCGCAA 2272

Qy       2182 CAGGAAATAATGGGAGTCTTCAACAAGAATTGCATGATCACATACGCCGCATACAGAAAC 2241

Db       2273 GAGGAAATAATGGGAGTCTTCAACAAGAATTGCATGATCACTTATGCTGCATACAGATGC 2332

Qy       2242 ATATTCCCCATTTGGGCGTTGGGAGAATACCAATCCCAAGTATTGCAATCTCGTTAATCG 2301

Db       2333 ATATTCCCCATATGGGCGTTAGGAGAATACC---GCCGTGTATTGCAGGCCTGCTAATAT 2389

Qy       2302 AGCCTTAGTTAGGGTGTCATCCTAATTATTTCGACCTGT 2340

Db       2390 CGCTTTATTTAGGCCACCATTCAAATTATTTCGACCTTT 2428


RESULT 4
AB055510
LOCUS       AB055510               2514 bp    mRNA    linear   PLN 05-JAN-2002
DEFINITION  Betula platyphylla CASBPX2 mRNA for cycloartenol synthase, complete
            cds.
ACCESSION   AB055510
VERSION     AB055510.1  GI:18147591
KEYWORDS    .
SOURCE      Betula platyphylla
  ORGANISM  Betula platyphylla
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fagales; Betulaceae; Betula.
REFERENCE   1
  AUTHORS   Zhang,H., Shibuya,M., Ebizuka,Y. and Yokota,S.
  TITLE     Oxidosqualene Cyclases from Cell Suspension Culture of Betula
            platyphylla var. japonica:Molecular Evolution of Oxidosqualene
            Cyclases in Higher Plants
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2514)
  AUTHORS   Shibuya,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-FEB-2001) Masaaki Shibuya, University of Tokyo,
            Faculty of Pharmaceutical Sciences; 7-3-1 Hongo, Bunkyo-ku, Tokyo
            113-0033, Japan (E-mail:shibuyam@mol.f.u-tokyo.ac.jp,
            URL:http://www.f.u-tokyo.ac.jp/~tennen/head.htm,
            Tel:81-3-5841-4741; Fax:81-3-5841-4744)
FEATURES             Location/Qualifiers
     source          1. .2514
                     /organism="Betula platyphylla"
                     /mol_type="mRNA"
                     /db_xref="taxon:78630"
     gene            1. .2514
                     /gene="CASBPX2"
     CDS             162. .2435
                     /gene="CASBPX2"
                     /EC_number="5.4.99.8"
                     /codon_start=1
                     /product="cycloartenol synthase"
                     /protein_id="BAB83086.1"
                     /db_xref="GI:18147592"
                     /translation="MWKLKIAEGGSPWLRTLNNHVGRQVWEFDPKLGSPEELAEIERA
                     RETSLKVRFEKKHSSDLLMRIQFAKENPRGAVLPQVKVNETEDVTEEMVTRMLRRAIS
                     FHSTLQAHDQHWAGDYGGPMFLMPGLVITLSITGALNTVLSEEHKKEMCRYLYNHQNK
                     DGGWGLHIEGPSTMFGTVLSYVTLRLLGEGANDGQGAIERGRKWILDHGSATAIISWG
                     KMWLSVLGAFEWSGNNPLPPEIWLLPYMLPVHPGRMWCHCRMVYLPMSYLYGKRFVGP
                     ITPTVMSLRKELYSVPYHEIDWNQARNLCAKEXLYYPHPLVQDILWASLHKLVEPVLM
                     RWPGKRLREKALRTVLEHIHYEDENTRYICIGPVNKVLNMLCCWVEDPNSEAFKLHLP
                     RINDYLWIAEDGMKMQGYNGSQLWDTAFAVQAIISTNLFEEYGPTLEKAHMYIKKSQV
                     REDCPGDLDFWYRHISKGAWPFSTADHGWPISDCTAEGLKAALLLSKIPPDVVGEPLV
                     EERLYDAVNVILSLQNADGGFATYELTRSYPWLELINPAETFGDIVIDYNYVECTSAA
                     IQALTSFKKSYPKHREEEVDVCIKRAAMFTEKIQASDGSWYGSWGVCFTYGTWFGVKG
                     LVAAGKNFNDCFGIRKACDFLLSKQLPSGGWGESYLSCQNKVYSHVEGNRSHVVNTGW
                     AMLALIEAGQAERDPTPLHRAARVLINSQMENGDFPQEEIMGVFNRNCMITYAAYRNI
                     FPIWALGEYRCRVLQAP"
ORIGIN

  Query Match           62.6%;  Score 1603.6;  DB 8;  Length 2514;
  Best Local Similarity  81.0%;  Pred. No. 0;
  Matches 1864;  Conservative    1;  Mismatches  435;  Indels    0;  Gaps    0;

Qy          5 CCTCTTGCCAGCAAAACAGAATGTGGAAGCTCAAGTTCGCCGAAGGAGGGAATCCATGGC 64

Db        142 CATTGTGCGAATAGTAGACAATGTGGAAGCTGAAGATCGCGGAGGGAGGGAGTCCATGGC 201

Qy         65 TTCGGACATTGAACAATCACGTTGGAAGACAGGTGTGGGAGTTCGATCCTAAGCTTGGAT 124

Db        202 TGCGGACGTTGAACAATCACGTCGGACGACAAGTTTGGGAATTCGATCCCAAGCTCGGGT 261

Qy        125 CGCCGCAAGATCTCCTCGAGATTGAGAAAGCTCGCCAGAATTTTCACGATAACCGCTTTA 184

Db        262 CGCCTGAAGAGCTTGCCGAGATCGAAAGAGCTCGCGAGACTTCTCTGAAAGTCCGTTTCG 321

Qy        185 CCCACAAACACAGCGCTGATCTACTTATGCGGATGCAGTTCGCAAGAGAGAACCCAACAC 244

Db        322 AGAAAAAGCATAGCTCAGATCTACTTATGCGGATTCAGTTCGCCAAGGAGAACCCACGTG 381

Qy        245 GTGAAGTCTTGCCCAAAGTCGGAGTTAAGGATATTGAGGATGTGACCCAAGAGATTGTGA 304

Db        382 GTGCAGTTCTACCCCAAGTCAAAGTAAATGAAACAGAAGATGTTACAGAGGAGATGGTGA 441

Qy        305 CAAAAACATTAAGAAGGGCCGTAAGTTTCCATTCAACTCTCCAGTGCCATGACGGACACT 364

Db        442 CAAGGATGTTAAGAAGGGCTATAAGTTTCCACTCAACTCTCCAGGCCCATGATGGACATT 501

Qy        365 GGCCGGGAGATTATGGAGGTCCCATGTTTCTGATGCCTGGCTTGGTAATTACTCTGTCTA 424

Db        502 GGGCGGGGGACTATGGAGGTCCAATGTTTCTTATGCCTGGCTTGGTAATCACTCTGTCTA 561

Qy        425 TCACTGGGGCGTTGAATACAGTCTTAACTGAAGAACATAGAAAGGAAATATGCCGTTACC 484

Db        562 TCACTGGAGCACTAAATACAGTTTATCTGAAGAACATAAAAAGGAAATGTGCCGTTACC 621

Qy        485 TCTATAATCATCAAAACAAGGATGGTGGGTGGGGTTTGCATATTGAAGGTCCAAGCACCA 544

Db        622 TCTATAATCATCAGAACAAAGATGGTGGGTGGGGATTACACATTGAGGGCCCAAGTACCA 681

Qy        545 TGTTTGGCTCTGTCTTGAGTTATATTACTCTGAGATTGCTAGGTGAGGGGCCTAATGATG 604

Db        682 TGTTTGGTACTGTCTTAAGCTATGTTACGTTAAGGTTGCTTGGTGAAGGAGCTAATGATG 741

Qy        605 GACAAGGGGAAATGGAGAAGGCACGTGACTGGATTCTAGGGCATGGTGGTGCTACTTATA 664

Db        742 GGCAAGGAGCAATTGAGAGAGGGCGTAAATGGATTCTGGACCATGGCAGTGCTACTGCAA 801

Qy        665 TAACGTCATGGGGGAAGATGTGGCTTTCAGTACTTGGAGTGTATGAATGGTCTGGAAATA 724

Db        802 TTATATCATGGGGAAAGATGTGGCTTTCAGTACTTGGAGCATTTGAATGGTCGGGCAATA 861

Qy        725 ATCCCCTGCCCCCTGAGATATGGCTCCTTCCATACATGCTTCCATTTCATCCAGGAAGGA 784

Db        862 ATCCCCTTCCCCCTGAAATATGGCTTCTTCCTTACATGCTCCCTGTACATCCAGGAAGAA 921

Qy        785 TGTGGTGTCACTGCCGGATGGTCTATTTGCCGATGTCCTACTTATATGGCAAGAGGTTTG 844

Db        922 TGTGGTGTCACTGTCGGATGGTCTATTTGCCAATGTCTTACTTATATGGGAAGAGATTTG 981

Qy        845 TTGGTCCAATCTCACCAACAGTATTATCTTTGAGAAAAGAGCTTTATACAGTACCATACC 904
```

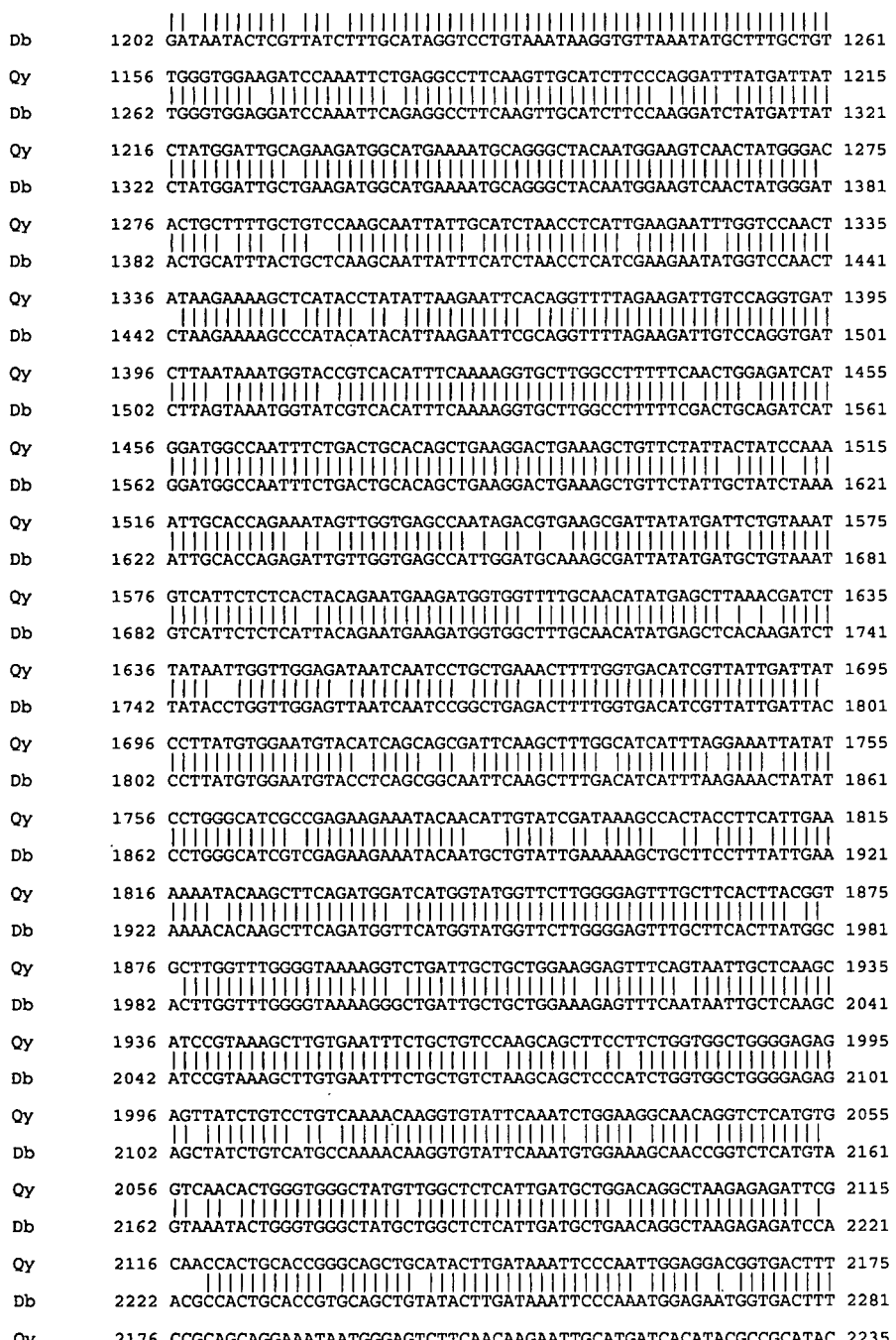
KMWLSVLGVFEWSGNNPMPPEIWLLPYALPVHPGRMWCHCRMVYLPMSYLYGKRFVGP
ITPTVLSLRKELFTVPYHDIDWNQARNLCAKEDLYYPHPLVQDILWATLHKFVEPVFM
NWPGKKLREKAIKTAIEHIHYEDENTRYICIGPVNKVLNMLCCWVEDPNSEAFKLHLP
RIYDYLWVAEDGMKMQGYNGSQLWDTAFAAQAIISTNLIDEFGPTLKKAHAFIKNSQV
SEDCPGDLSKWYRHISKGAWPFSTADHGWPISDCTAEGLKAVLLLSKIAPEIVGEPLD
SKRLYDAVNVILSLQNENGGLATYELTRSYTWLEIINPAETFGDIVIDCPYVECTSAA
IQALATFGKLYPGHRREEIQCCIEKAVAFIEKIQASDGSWYGSWGVCFTYGTWFGIKG
LIAAGKNFSNCLSIRKACEFLLSKQLPSGGWAESYLSCQNKVYSNLEGNRSHVVNTGW
AMLALIEAEQAKRDPTPLHRAAVCLINSQLENGDFPQEEIMGVFNKNCMITYAAYRCI
FPIWALGEYRRVLQAC"

ORIGIN

Query Match 71.6%; Score 1834.2; DB 8; Length 2677;
Best Local Similarity 87.2%; Pred. No. 0;
Matches 2023; Conservative 0; Mismatches 293; Indels 3; Gaps 1;

```
Qy      22 AGAATGTGGAAGCTCAAGTTCGCCGAAGGAGGGAATCCATGGCTTCGGACATTGAACAAT 81
Db     113 AAAATGTGGAAGCTCAAGGTTGCGGAAGGAGGTACTCCATGGCTTCGGACATTGAACAAC 172
Qy      82 CACGTTGGAAGACAGGTGTGGGAGTTCGATCCTAAGCTTGGATCGCCGCAAGATCTCCTC 141
Db     173 CACGTCGGAAGACAGGTATGGGAGTTCGATCCTCATTCCGGATCGCCACAAGATCTCGAC 232
Qy     142 GAGATTGAGAAAGCTCGCCAGAATTTTCACGATAACCGCTTTACCCACAAACACAGCGCT 201
Db     233 GATATCGAAACCGCTCGCCGCAATTTCCACGATAATCGCTTCACTCACAAACACAGTGAT 292
Qy     202 GATCTACTTATGCGGATGCAGTTCGCAAGAGAGAACCCAACACGTGAAGTCTTGCCCAAA 261
Db     293 GATCTACTTATGCGCCTTCAGTTTGCGAAAGAGAACCCTATGAACGAAGTTTTACCTAAA 352
Qy     262 GTCGGAGTTAAGGATATTGAGGATGTGACCCAAGAGATTGTGACAAAAACATTAAGAAGG 321
Db     353 GTGAAAGTCAAAGATGTCGAGGATGTAACGGAAGAGGCTGTTGCAACGACATTAAGAAGG 412
Qy     322 GCCGTAAGTTTCCATTCAACTCTCCAGTGCCATGACGGACACTGGCCGGGAGATTATGGA 381
Db     413 GGATTGAATTTCTATTCAACTATCCAGAGTCATGATGGACACTGGCCAGGAGATTATGGA 472
Qy     382 GGTCCCATGTTTCTGATGCCTGGCTTGGTAATTACTCTGTCTATCACTGGGGCGTTGAAT 441
Db     473 GGTCCCATGTTTCTTATGCCTGGATTGGTTATTACTCTGTCTGTCACCGGGGCTTTGAAT 532
Qy     442 ACAGTCTTAACTGAAGAACATAGAAAGGAAATATGCCGTTACCTCTATAATCATCAAAAC 501
Db     533 GCAGTCTTAACAGACGAGCATAGAAAGGAAATGCGTCGTTACCTTTATAATCATCAGAAC 592
Qy     502 AAGGATGGTGGGTGGGGTTTGCATATTGAAGGTCCAAGCACCATGTTTGGCTCTGTCTTG 561
Db     593 AAGGATGGTGGGTGGGGTTTGCATATTGAGGGTCCAAGCACAATGTTTGGCTCTGTCTTG 652
Qy     562 AGTTATATTACTCTGAGATTGCTAGGTGAGGGGCCTAATGATGGACAAGGGGAAATGGAG 621
Db     653 TGTTATGTTACTCTAAGATTGCTAGGGGAGGGACCTAATGATGGAGAAGGGGATATGGAG 712
Qy     622 AAGGCACGTGACTGGATTCTAGGGCATGGTGGTGCTACTTATATAACGTCATGGGGGAAG 681
Db     713 AGGGGGCGTGATTGGATTCTTGAGCATGGTGGTGCTACTTATATAACATCCTGGGGGAAA 772
Qy     682 ATGTGGCTTTCAGTACTTGGAGTGTATGAATGGTCTGGAAATAATCCCCTGCCCCCTGAG 741
Db     773 ATGTGGCTATCAGTACTTGGAGTGTTTGAGTGGTCTGGAAATAATCCCATGCCTCCTGAA 832
Qy     742 ATATGGCTCCTTCCATACATGCTTCCATTTCATCCAGGAAGGATGTGGTGTCACTGCCGG 801
Db     833 ATATGGCTCCTTCCATACGCGCTTCCTGTTCATCCAGGAAGGATGTGGTGTCACTGCCGG 892
Qy     802 ATGGTCTATTTGCCGATGTCCTACTTATATGGCAAGAGGTTTGTTGGTCCAATCTCACCA 861
Db     893 ATGGTCTATTTGCCAATGTCCTACTTGTATGGCAAGAGGTTTGTTGGTCCAATCACACCA 952
Qy     862 ACAGTATTATCTTTGAGAAAAGAGCTTTATACAGTACCATACCATGATATAGATTGGGAT 921

Db     953 ACAGTATTATCTTTGAGAAAAGAGCTTTTTACCGTTCCATACCATGATATAGACTGGAAC 1012
Qy     922 CAGGCTCGCAATTTGTGTGCAAAGGAAGATTTGTACTATCCTCACCCACTTGTACAGGAT 981
Db    1013 CAGGCTCGCAATTTGTGTGCAAAGGAAGACTTATATTATCCTCATCCGCTTGTGCAAGAT 1072
Qy     982 ATTCTTTGGGCATCTCTACACAAGTTCCTTGAGCCTATTCTGATGCATTGGCCTGGAAAA 1041
Db    1073 ATTCTTTGGGCAACTCTACACAAGTTTGTTGAGCCTGTTTTCATGAATTGGCCTGGCAAA 1132
Qy    1042 AGATTGAGGGAAAAGGCTATTATTTCTGCATTGGAGCATATACATTACGAAGATGAGAAT 1101
Db    1133 AAGTTGAGAGAAAAGGCTATCAAGACTGCAATTGAGCATATACATTATGAAGATGAGAAC 1192
Qy    1102 ACTCGATATATTTGCATAGGTCCTGTAAATAAGGTGTTAAATATGCTTTGCTGTTGGGTG 1161
Db    1193 ACTCGTTATATTTGCATAGGTCCTGTAAATAAGGTGTTAAATATGCTTTGCTGCTGGGTA 1252
Qy    1162 GAAGATCCAAATTCTGAGGCCTTCAAGTTGCATCTTCCCAGGATTTATGATTATCTATGG 1221
Db    1253 GAAGATCCAAATTCAGAGGCTTTCAAGTTGCATCTTCCCAGGATCTATGATTATCTATGG 1312
Qy    1222 ATTGCAGAAGATGGCATGAAAATGCAGGGCTACAATGGAAGTCAACTATGGGACACTGCT 1281
Db    1313 GTTGCTGAAGATGGCATGAAAATGCAGGGCTACAATGGAAGTCAACTATGGGATACTGCA 1372
Qy    1282 TTTGCTGTCCAAGCAATTATTGCATCTAACCTCATTGAAGAATTTGGTCCAACTATAAGA 1341
Db    1373 TTTGCTGCACAAGCAATTATTTCAACTAACGTAATTGACGAATTTGGTCCAACTCTAAAA 1432
Qy    1342 AAAGCTCATACCTATATTAAGAATTCACAGGTTTTAGAAGATTGTCCAGGTGATCTTAAT 1401
Db    1433 AAGGCTCATGCATTCATTAAGAATTCACAGGTTTCGGAAGATTGTCCGGGGGATCTTAGT 1492
Qy    1402 AAATGGTACCGTCACATTTCAAAAGGTGCTTGGCCTTTTTCAACTGGAGATCATGGATGG 1461
Db    1493 AAATGGTACCGTCACATTTCAAAAGGTGCTTGGCCTTTTTCAACTGCAGACCATGGATGG 1552
Qy    1462 CCAATTTCTGACTGCACAGCTGAAGGACTGAAAGCTGTTCTATTACTATCCAAAATTGCA 1521
Db    1553 CCAATATCTGACTGCACGGCAGAAGGACTGAAAGCTGTTCTATTGCTATCTAAAATTGCA 1612
Qy    1522 CCAGAAATAGTTGGTGAGCCAATAGACGTGAAGCGATTATATGATTCTGTAAATGTCATT 1581
Db    1613 CCTGAGATTGTTGGTGAACCATTGGACTCAAAGAGATTATATGATGCTGTAAATGTTATT 1672
Qy    1582 CTCTCACTACAGAATGAAGATGGTGGTTTTGCAACATATGAGCTTAAACGATCTTATAAT 1641
Db    1673 CTCTCATTACAGAATGAAAATGGTGGCCTTGCAACATATGAGCTCACACGATCTTATACC 1732
Qy    1642 TGGTTGGAGATAATCAATCCTGCTGAAACTTTGGTGACATCGTTATTGATTATCCTTAT 1701
Db    1733 TGGTTGGAGATAATCAATCCTGCTGAAACATTTGGTGACATTGTTATTGATTGTCCTTAC 1792
Qy    1702 GTGGAATGTACATCAGCAGCGATTCAAGCTTTGGCATCATTTAGGAAATTATATCCTGGG 1761
Db    1793 GTGGAATGTACCTCAGCAGCAATTCAAGCTTTAGCAACATTTGGGAAACTATATCCCGGG 1852
Qy    1762 CATCGCCGAGAAGAAATACAACATTGTATCGATAAAGCCACTACCTTCATTGAAAAAATA 1821
Db    1853 CATCGTCGAGAAGAAATACAATGTTGTATAGAAAAAGCTGTTGCCTTTATTGAAAAAATT 1912
Qy    1822 CAAGCTTCAGATGGATCATGGTATGGTTCTTGGGGAGTTTGCTTCACTTACGGTGCTTGG 1881
Db    1913 CAAGCTTCAGATGGTTCATGGTATGGTTCTTGGGGAGTTTGCTTTACATATGGCACTTGG 1972
Qy    1882 TTTGGGGTAAAAGGTCTGATTGCTGCTGGAAGGAGTTTCAGTAATTGCTCAAGCATCCGT 1941
Db    1973 TTTGGGATAAAAGGGCTGATTGCTGCTGGAAAGAATTTCAGTAATTGTTTAAGCATCCGT 2032
Qy    1942 AAAGCTTGTGAATTTCTGCTGTCCAAGCAGCTTCCTTCTGGTGGCTGGGGAGAGAGTTAT 2001
```



RESULT 3
D89619
LOCUS D89619 2677 bp mRNA linear PLN 14-JAN-2000
DEFINITION Pisum sativum mRNA for cycloartenol synthase, complete cds.
ACCESSION D89619
VERSION D89619.1 GI:2627180
KEYWORDS cycloartenol synthase.
SOURCE Pisum sativum (pea)
ORGANISM Pisum sativum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Vicieae; Pisum.
REFERENCE 1 (sites)
AUTHORS Morita,M., Shibuya,M., Lee,M.S., Sankawa,U. and Ebizuka,Y.
TITLE Molecular cloning of pea cDNA encoding cycloartenol synthase and its functional expression in yeast
JOURNAL Biol. Pharm. Bull. 20 (7), 770-775 (1997)
MEDLINE 97399291
PUBMED 9255418
REFERENCE 2 (bases 1 to 2677)
AUTHORS Shibuya,M.
TITLE Direct Submission
JOURNAL Submitted (28-NOV-1996) Masaaki Shibuya, The University of Tokyo, Faculty of Pharmaceutical Sciences; 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan (E-mail:shibuyam@mol.f.u-tokyo.ac.jp, Tel:81-3-5841-4741, Fax:81-3-5841-4744)
FEATURES Location/Qualifiers
source 1. .2677
/organism="Pisum sativum"
/mol_type="mRNA"
/db_xref="taxon:3888"
/dev_stage="germination"
gene 1. .2677
/gene="CASPEA"
CDS 116. .2386
/gene="CASPEA"
/EC_number="5.4.99.8"
/codon_start=1
/product="cycloartenol synthase"
/protein_id="BAA23533.1"
/db_xref="GI:2627181"
/translation="MWKLKVAEGGTPWLRTLNNHVGRQVWEFDPHSGSPQDLDDIETA
RRNFHDNRFTHKHSDDLLMRLQFAKENPMNEVLPKVKVKDVEDVTEEAVATTLRRGLN
FYSTIQSHDGHWPGDYGGPMFLMPGLVITLSVTGALNAVLTDEHRKEMRRYLYNHQNK
DGGWGLHIEGPSTMFGSVLCYVTLRLLGEGPNDGEGDMERGRDWILEHGGATYITSWG

```
Qy       2461 GCCTGTAAGTTTTTTCTATTACTCATAGACGCCTCTCTTGCTTTCCCTGCAGCAGAGGAA 2520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2461 GCCTGTAAGTTTTTTCTATTACTCATAGACGCCTCTCTTGCTTTCCCTGCAGCAGAGGAA 2520

Qy       2521 TTAAATATATACATAAATAGAGATATAAAAAAAAAAAAAAA 2560
             ||||||||||||||||||||||||||||||||||||||||||
Db       2521 TTAAATATATACATAAATAGAGATATAAAAAAAAAAAAAA 2560


RESULT 2
AB025968
LOCUS       AB025968               2889 bp    mRNA    linear   PLN 14-DEC-2001
DEFINITION  Glycyrrhiza glabra mRNA for cycloartenol synthase, complete cds.
ACCESSION   AB025968
VERSION     AB025968.1  GI:4589851
KEYWORDS    cycloartenol synthase.
SOURCE      Glycyrrhiza glabra
  ORGANISM  Glycyrrhiza glabra
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Galegeae;
            Glycyrrhiza.
REFERENCE   1
  AUTHORS   Hayashi,H., Hiraoka,N., Ikeshiro,Y., Kushiro,T., Morita,M.,
            Shibuya,M. and Ebizuka,Y.
  TITLE     Molecular cloning and characterization of a cDNA for Glycyrrhiza
            glabra cycloartenol synthase
  JOURNAL   Biol. Pharm. Bull. 23 (2), 231-234 (2000)
  MEDLINE   20168908
   PUBMED   10706391
REFERENCE   2  (bases 1 to 2889)
  AUTHORS   Hayashi,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-APR-1999) Hiroaki Hayashi, Gifu Pharmaceutical
            University, Department of Pharmacognosy; 5-6-1 Mitahora-higashi,
            Gifu, Gifu 502-8585, Japan (E-mail:hayashih@gifu-pu.ac.jp,
            Tel:81-58-237-3931(ex.325), Fax:81-58-237-5979)
FEATURES             Location/Qualifiers
     source          1. .2889
                     /organism="Glycyrrhiza glabra"
                     /mol_type="mRNA"
                     /db_xref="taxon:49827"
                     /clone="GgCAS1"
                     /cell_type="suspension cultured cells"
                     /clone_lib="lambda ZAPII"
     gene            1. .2889
                     /gene="GgCAS1"
     CDS             131. .2404
                     /gene="GgCAS1"
                     /codon_start=1
                     /product="cycloartenol synthase"
                     /protein_id="BAA76902.1"
                     /db_xref="GI:4589852"
                     /translation="MWKLKIAEGGSPWLRTVNNHVGRQVWEFDPKLGSPEDLLEIEKA
                     RQNFHDNRFTHKHSADLLMRIHFAKENPMNEVLPKVRVKDIEDVTEETVKTTLRRAIN
                     FHSTLQSHDGHWPGDYGGPMFLMPGLVITLSITGALNAVLTEEHRKEICRYLYNHQNK
                     DGGWGLHIEGPSTMFGSVLNYVALRLLGEGPNDRQGEMEKGRDWILGHGGATFITSWG
                     KMWLSVLGVYEWSGNNPLPPEIWLLPYVLPIHPGRMWCHCRMVYLPMSYLYGKRFVGP
                     ITPTILSLRKELYTIPYHDIDWNQARNLCAKEDLYYPHPLVQDILWASLHKFLEPILM
                     HWPGKKLREMAIKTAIEHIHYEDDNTRYLCIGPVNKVLNMLCCWVEDPNSEAFKLHLP
                     RIYDYLWIAEDGMKMQGYNGSQLWDTAFTAQAIISSNLIEEYGPTLRKAHTYIKNSQV
                     LEDCPGDLSKWYRHISKGAWPFSTADHGWPISDCTAEGLKAVLLLSKIAPEIVGEPLD
                     AKRLYDAVNVILSLQNEDGGFATYELTRSYTWLELINPAETFGDIVIDYPYVECTSAA
                     IQALTSFKKLYPGHRREEIQCCIEKAASFIEKTQASDGSWYGSWGVCFTYGTWFGVKG
                     LIAAGKSFNNCSSIRKACEFLLSKQLPSGGWGESYLSCQNKVYSNVESNRSHVVNTGW
                     AMLALIDAEQAKRDPTPLHRAAVYLINSQMENGDFPQQEIMGVFNKNCMITYAAYRNV
                     FPIWALGEYRHRVLQSQ"
ORIGIN

  Query Match           77.1%;  Score 1973.8;  DB 8;  Length 2889;
  Best Local Similarity  87.1%;  Pred. No. 0;
  Matches 2205;  Conservative    0;  Mismatches  292;  Indels   36;  Gaps    2;
```

```
Qy         16 CAAAACAGAATGTGGAAGCTCAAGTTCGCCGAAGGAGGGAATCCATGGCTTCGGACATTG 75
              ||  |  |||||||||||||||| | || || ||||||| ||||||||||||||||| ||
Db        122 CAGAGGAAAATGTGGAAGCTCAAGATTGCGGAGGGAGGGAGTCCATGGCTTCGGACAGTG 181

Qy         76 AACAATCACGTTGGAAGACAGGTGTGGGAGTTCGATCCTAAGCTTGGATCGCCGCAAGAT 135
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 AACAATCACGTTGGAAGACAGGTGTGGGAGTTCGATCCTAAGCTTGGATCGCCAGAAGAT 241

Qy        136 CTCCTCGAGATTGAGAAAGCTCGCCAGAATTTTCACGATAACCGCTTTACCCACAAACAC 195
              |||||||||||||| || |||||||||||||| ||||| |||||||| || |||||||||
Db        242 CTCCTCGAGATTGAAAAGGCTCGCCAGAATTTCCACGACAACCGCTTCACTCACAAACAC 301

Qy        196 AGCGCTGATCTACTTATGCGGATGCAGTTCGCAAGAGAGAACCCAACACGTGAAGTCTTG 255
              |||||||||||||| ||||| || || || |||| ||||||||||     |||||| |
Db        302 AGCGCTGATCTACTCATGCGCATTCACTTTGCAAAAGAGAACCCAATGAACGAAGTCCTA 361

Qy        256 CCCAAAGTCGGAGTTAAGGATATTGAGGATGTGACCCAAGAGATTGTGACAAAAACATTA 315
              |||||||| |||| || |||||||| ||||| || |||||| |||| || ||||||
Db        362 CCCAAAGTCAGAGTCAAAGATATTGAAGATGTAACTGAAGAGACCGTGAAAACTACATTA 421

Qy        316 AGAAGGGCCGTAAGTTTCCATTCAACTCTCCAGTGCCATGACGGACACTGGCCGGGAGAT 375
              ||||||||| ||| |||||||||||||||| | ||||| |||||||||| ||||||
Db        422 AGAAGGGCCATAAATTTCCATTCAACTCTCCAGAGTCATGATGGACACTGGCCAGGAGAT 481

Qy        376 TATGGAGGTCCCATGTTTCTGATGCCTGGCTTGGTAATTACTCTGTCTATCACTGGGGCG 435
              |||||||||||||||||| |||||||||||||||||||||||||||||||||| |||||
Db        482 TATGGAGGTCCCATGTTTCTTATGCCTGGCTTGGTAATTACTCTGTCTATCACCGGGGCT 541

Qy        436 TTGAATACAGTCTTAACTGAAGAACATAGAAAGGAAATATGCCGTTACCTCTATAATCAT 495
              ||||| |||||||||| ||||| |||||||||||||||||||||||||||||||| |||
Db        542 CTGAATGCAGTCTTAACAGAAGAGCATAGAAAGGAAATATGCCGTTACCTCTATAACCAT 601

Qy        496 CAAAACAAGGATGGTGGGTGGGGTTTGCATATTGAAGGTCCAAGCACCATGTTTGGCTCT 555
              || |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        602 CAGAACAAGGATGGTGGGTGGGGTTTACATATTGAAGGTCCAAGCACCATGTTTGGCTCT 661

Qy        556 GTCTTGAGTTATATTACTCTGAGATTGCTAGGTGAGGGGCCTAATGATGGACAAGGGGAA 615
              ||||||| |||| || |||||||||||| || ||||| |||||||||| ||||||||||
Db        662 GTCTTGAATTATGTTGCTCTGAGATTGCTTGGCGAGGGACCTAATGATAGACAAGGGGAA 721

Qy        616 ATGGAGAAGGCACGTGACTGGATTCTAGGGCATGGTGGTGCTACTTATATAACGTCATGG 675
              |||||||||| |||||||||||||| |||||||| |||||||||| |||||| || |||
Db        722 ATGGAGAAGGGACGTGACTGGATTCTTGGGCATGGCGGTGCTACTTTTATAACATCGTGG 781

Qy        676 GGGAAGATGTGGCTTTCAGTACTTGGAGTGTATGAATGGTCTGGAAATAATCCCCTGCCC 735
              || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        782 GGAAAGATGTGGCTTTCAGTACTTGGAGTGTATGAATGGTCTGGAAATAATCCCCTGCCT 841

Qy        736 CCTGAGATATGGCTCCTTCCATACATGCTTCCATTTCATCCAGGAAGGATGTGGTGTCAC 795
              |||||||||||||||||||||| | |||||| ||||||||||||||||||||||||||
Db        842 CCTGAGATATGGCTCCTTCCATACGTTCTTCCAATTCATCCAGGAAGGATGTGGTGTCAC 901

Qy        796 TGCCGGATGGTCTATTTGCCGATGTCCTACTTATATGGCAAGAGGTTTGTTGGTCCAATC 855
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        902 TGCCGGATGGTCTATTTGCCAATGTCCTACTTATATGGAAAGAGGTTTGTTGGCCCAATC 961

Qy        856 TCACCAACAGTATTATCTTTGAGAAAAGAGCTTTATACAGTACCATACCATGATATAGAT 915
              |||||||| |||||||||||||||| |||||||||| |||||| ||||||||||
Db        962 ACACCAACAATATTATCTTTGAGAAAGGAGCTTTATACCATACCATATCATGATATAGAC 1021

Qy        916 TGGGATCAGGCTCGCAATTTGTGTGCAAAGGAAGATTTGTACTATCCTCACCCACTTGTA 975
              ||| |||||||||||||||||||||||||||||||| || ||||||||||||||||||||
Db       1022 TGGAATCAGGCTCGCAATTTGTGTGCAAAGGAAGACTTATACTATCCTCACCCACTTGTG 1081

Qy        976 CAGGATATTCTTTGGGCATCTCTACACAAGTTCCTTGAGCCTATTCTGATGCATTGGCCT 1035
              || |||||||| |||||||||||||||||||||||||| | |||||||||| ||||||||||
Db       1082 CAAGATATTCTGTGGGCATCTCTACACAAGTTCTTGGAGCCTATTTTGATGCATTGGCCT 1141

Qy       1036 GGAAAAAGATTGAGGGAAAAGGCTATTATTTCTGCATTGGAGCATATACATTACGAAGAT 1095
              || ||||  |||||  |||| |||||||   |||||  | ||||||||||||| ||||||
Db       1142 GGCAAAAAGTTGAGAGAAATGGCTATTAAGACTGCAATCGAGCATATACATTATGAAGAT 1201

Qy       1096 GAGAATACTCGATATATTTGCATAGGTCCTGTAAATAAGGTGTTAAATATGCTTTGCTGT 1155
```

```
Db        241 ACACGTGAAGTCTTGCCCAAAGTCGGAGTTAAGGATATTGAGGATGTGACCCAAGAGATT 300
Qy        301 GTGACAAAAACATTAAGAAGGGCCGTAAGTTTCCATTCAACTCTCCAGTGCCATGACGGA 360
Db        301 GTGACAAAAACATTAAGAAGGGCCGTAAGTTTCCATTCAACTCTCCAGTGCCATGACGGA 360
Qy        361 CACTGGCCGGGAGATTATGGAGGTCCCATGTTTCTGATGCCTGGCTTGGTAATTACTCTG 420
Db        361 CACTGGCCGGGAGATTATGGAGGTCCCATGTTTCTGATGCCTGGCTTGGTAATTACTCTG 420
Qy        421 TCTATCACTGGGGCGTTGAATACAGTCTTAACTGAAGAACATAGAAAGGAAATATGCCGT 480
Db        421 TCTATCACTGGGGCGTTGAATACAGTCTTAACTGAAGAACATAGAAAGGAAATATGCCGT 480
Qy        481 TACCTCTATAATCATCAAAACAAGGATGGTGGGTGGGGTTTGCATATTGAAGGTCCAAGC 540
Db        481 TACCTCTATAATCATCAAAACAAGGATGGTGGGTGGGGTTTGCATATTGAAGGTCCAAGC 540
Qy        541 ACCATGTTTGGCTCTGTCTTGAGTTATATTACTCTGAGATTGCTAGGTGAGGGGCCTAAT 600
Db        541 ACCATGTTTGGCTCTGTCTTGAGTTATATTACTCTGAGATTGCTAGGTGAGGGGCCTAAT 600
Qy        601 GATGGACAAGGGGAAATGGAGAAGGCACGTGACTGGATTCTAGGGCATGGTGGTGCTACT 660
Db        601 GATGGACAAGGGGAAATGGAGAAGGCACGTGACTGGATTCTAGGGCATGGTGGTGCTACT 660
Qy        661 TATATAACGTCATGGGGGAAGATGTGGCTTTCAGTACTTGGAGTGTATGAATGGTCTGGA 720
Db        661 TATATAACGTCATGGGGGAAGATGTGGCTTTCAGTACTTGGAGTGTATGAATGGTCTGGA 720
Qy        721 AATAATCCCCTGCCCCCTGAGATATGGCTCCTTCCATACATGCTTCCATTTCATCCAGGA 780
Db        721 AATAATCCCCTGCCCCCTGAGATATGGCTCCTTCCATACATGCTTCCATTTCATCCAGGA 780
Qy        781 AGGATGTGGTGTCACTGCCGGATGGTCTATTTGCCGATGTCCTACTTATATGGCAAGAGG 840
Db        781 AGGATGTGGTGTCACTGCCGGATGGTCTATTTGCCGATGTCCTACTTATATGGCAAGAGG 840
Qy        841 TTTGTTGGTCCAATCTCACCAACAGTATTATCTTTGAGAAAAGAGCTTTATACAGTACCA 900
Db        841 TTTGTTGGTCCAATCTCACCAACAGTATTATCTTTGAGAAAAGAGCTTTATACAGTACCA 900
Qy        901 TACCATGATATAGATTGGGATCAGGCTCGCAATTTGTGTGCAAAGGAAGATTTGTACTAT 960
Db        901 TACCATGATATAGATTGGGATCAGGCTCGCAATTTGTGTGCAAAGGAAGATTTGTACTAT 960
Qy        961 CCTCACCCACTTGTACAGGATATTCTTTGGGCATCTCTACACAAGTTCCTTGAGCCTATT 1020
Db        961 CCTCACCCACTTGTACAGGATATTCTTTGGGCATCTCTACACAAGTTCCTTGAGCCTATT 1020
Qy       1021 CTGATGCATTGGCCTGGAAAAAGATTGAGGGAAAAGGCTATTATTTCTGCATTGGAGCAT 1080
Db       1021 CTGATGCATTGGCCTGGAAAAAGATTGAGGGAAAAGGCTATTATTTCTGCATTGGAGCAT 1080
Qy       1081 ATACATTACGAAGATGAGAATACTCGATATATTTGCATAGGTCCTGTAAATAAGGTGTTA 1140
Db       1081 ATACATTACGAAGATGAGAATACTCGATATATTTGCATAGGTCCTGTAAATAAGGTGTTA 1140
Qy       1141 AATATGCTTTGCTGTTGGGTGGAAGATCCAAATTCTGAGGCCTTCAAGTTGCATCTTCCC 1200
Db       1141 AATATGCTTTGCTGTTGGGTGGAAGATCCAAATTCTGAGGCCTTCAAGTTGCATCTTCCC 1200
Qy       1201 AGGATTTATGATTATCTATGGATTGCAGAAGATGGCATGAAAATGCAGGGCTACAATGGA 1260
Db       1201 AGGATTTATGATTATCTATGGATTGCAGAAGATGGCATGAAAATGCAGGGCTACAATGGA 1260
Qy       1261 AGTCAACTATGGGACACTGCTTTTGCTGTCCAAGCAATTATTGCATCTAACCTCATTGAA 1320
Db       1261 AGTCAACTATGGGACACTGCTTTTGCTGTCCAAGCAATTATTGCATCTAACCTCATTGAA 1320
Qy       1321 GAATTTGGTCCAACTATAAGAAAAGCTCATACCTATATTAAGAATTCACAGGTTTTAGAA 1380
Db       1321 GAATTTGGTCCAACTATAAGAAAAGCTCATACCTATATTAAGAATTCACAGGTTTTAGAA 1380
```

```
Qy       1381 GATTGTCCAGGTGATCTTAATAAATGGTACCGTCACATTTCAAAAGGTGCTTGGCCTTTT 1440
Db       1381 GATTGTCCAGGTGATCTTAATAAATGGTACCGTCACATTTCAAAAGGTGCTTGGCCTTTT 1440
Qy       1441 TCAACTGGAGATCATGGATGGCCAATTTCTGACTGCACAGCTGAAGGACTGAAAGCTGTT 1500
Db       1441 TCAACTGGAGATCATGGATGGCCAATTTCTGACTGCACAGCTGAAGGACTGAAAGCTGTT 1500
Qy       1501 CTATTACTATCCAAAATTGCACCAGAAATAGTTGGTGAGCCAATAGACGTGAAGCGATTA 1560
Db       1501 CTATTACTATCCAAAATTGCACCAGAAATAGTTGGTGAGCCAATAGACGTGAAGCGATTA 1560
Qy       1561 TATGATTCTGTAAATGTCATTCTCTCACTACAGAATGAAGATGGTGGTTTTGCAACATAT 1620
Db       1561 TATGATTCTGTAAATGTCATTCTCTCACTACAGAATGAAGATGGTGGTTTTGCAACATAT 1620
Qy       1621 GAGCTTAAACGATCTTATAATTGGTTGGAGATAATCAATCCTGCTGAAACTTTTGGTGAC 1680
Db       1621 GAGCTTAAACGATCTTATAATTGGTTGGAGATAATCAATCCTGCTGAAACTTTTGGTGAC 1680
Qy       1681 ATCGTTATTGATTATCCTTATGTGGAATGTACATCAGCAGCGATTCAAGCTTTGGCATCA 1740
Db       1681 ATCGTTATTGATTATCCTTATGTGGAATGTACATCAGCAGCGATTCAAGCTTTGGCATCA 1740
Qy       1741 TTTAGGAAATTATATCCTGGGCATCGCCGAGAAGAAATACAACATTGTATCGATAAAGCC 1800
Db       1741 TTTAGGAAATTATATCCTGGGCATCGCCGAGAAGAAATACAACATTGTATCGATAAAGCC 1800
Qy       1801 ACTACCTTCATTGAAAAAATACAAGCTTCAGATGGATCATGGTATGGTTCTTGGGGAGTT 1860
Db       1801 ACTACCTTCATTGAAAAAATACAAGCTTCAGATGGATCATGGTATGGTTCTTGGGGAGTT 1860
Qy       1861 TGCTTCACTTACGGTGCTTGGTTTGGGGTAAAAGGTCTGATTGCTGCTGGAAGGAGTTTC 1920
Db       1861 TGCTTCACTTACGGTGCTTGGTTTGGGGTAAAAGGTCTGATTGCTGCTGGAAGGAGTTTC 1920
Qy       1921 AGTAATTGCTCAAGCATCCGTAAAGCTTGTGAATTTCTGCTGTCCAAGCAGCTTCCTTCT 1980
Db       1921 AGTAATTGCTCAAGCATCCGTAAAGCTTGTGAATTTCTGCTGTCCAAGCAGCTTCCTTCT 1980
Qy       1981 GGTGGCTGGGGAGAGAGTTATCTGTCCTGTCAAAACAAGGTGTATTCAAATCTGGAAGGC 2040
Db       1981 GGTGGCTGGGGAGAGAGTTATCTGTCCTGTCAAAACAAGGTGTATTCAAATCTGGAAGGC 2040
Qy       2041 AACAGGTCTCATGTGGTCAACACTGGGTGGGCTATGTTGGCTCTCATTGATGCTGGACAG 2100
Db       2041 AACAGGTCTCATGTGGTCAACACTGGGTGGGCTATGTTGGCTCTCATTGATGCTGGACAG 2100
Qy       2101 GCTAAGAGAGATTCGCAACCACTGCACCGGGCAGCTGCATACTTGATAAATTCCCAATTG 2160
Db       2101 GCTAAGAGAGATTCGCAACCACTGCACCGGGCAGCTGCATACTTGATAAATTCCCAATTG 2160
Qy       2161 GAGGACGGTGACTTTCCGCAGCAGGAAATAATGGGAGTCTTCAACAAGAATTGCATGATC 2220
Db       2161 GAGGACGGTGACTTTCCGCAGCAGGAAATAATGGGAGTCTTCAACAAGAATTGCATGATC 2220
Qy       2221 ACATACGCCGCATACAGAAACATATTCCCCATTTGGGCGTTGGGAGAATACCAATCCCAA 2280
Db       2221 ACATACGCCGCATACAGAAACATATTCCCCATTTGGGCGTTGGGAGAATACCAATCCCAA 2280
Qy       2281 GTATTGCAATCTCGTTAATCGAGCCTTAGTTAGGGTGTCATCCTAATTATTTCGACCTGT 2340
Db       2281 GTATTGCAATCTCGTTAATCGAGCCTTAGTTAGGGTGTCATCCTAATTATTTCGACCTGT 2340
Qy       2341 GGACAGTAAAAGAAATAATAATAATAACCTATTTTTCTTTTATTTCCATGGCTCTCTTAA 2400
Db       2341 GGACAGTAAAAGAAATAATAATAATAACCTATTTTTCTTTTATTTCCATGGCTCTCTTAA 2400
Qy       2401 AATGTTTGTGACTAATGAGTTTAGTAGTCAGCTAAAAAAAAAAAGCAAACACGTGGAGAAT 2460
Db       2401 AATGTTTGTGACTAATGAGTTTAGTAGTCAGCTAAAAAAAAAAAGCAAACACGTGGAGAAT 2460
```

Baum, S.
09/419679
PN 6630617
Seq. ID 5 & 6

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2005, 06:25:28 ; Search time 10816 Seconds
(without alignments)
11468.690 Million cell updates/sec

Title: US-09-419-679-5
Perfect score: 2560
Sequence: 1 ttggcctcttgccagcaaaa.....agatataaaaaaaaaaaaaa 2560

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2560	100.0	2560	6	AR407878	AR407878 Sequence
2	1973.8	77.1	2889	8	AB025968	AB025968 Glycyrrhi
3	1834.2	71.6	2677	8	D89619	D89619 Pisum sativ
4	1603.6	62.6	2514	8	AB055510	AB055510 Betula pl
5	1502.4	58.7	2559	8	AB009029	AB009029 Panax gin
6	1501.2	58.6	2547	8	AY520819	AY520819 Centella
7	1411.4	55.1	2561	8	U02555	U02555 Arabidopsis
8	1408.2	55.0	2654	8	AY094394	AY094394 Arabidops
9	1405	54.9	2280	6	AX506138	AX506138 Sequence
10	1405	54.9	2280	8	BT001118	BT001118 Arabidops
11	1394	54.5	2460	8	AB055509	AB055509 Betula pl
12	1352.4	52.8	2556	8	AB116237	AB116237 Cucurbita
13	1351	52.8	2433	8	AB033334	AB033334 Luffa cyl
14	1306.2	51.0	2683	8	AB058507	AB058507 Costus sp
15	1226.6	47.9	2280	8	AY618689	AY618689 Avena cla
16	1223.6	47.8	2651	8	AK121211	AK121211 Oryza sat
17	1222	47.7	2413	8	AF169966	AF169966 Oryza sat
18	1218.6	47.6	2280	8	AY618690	AY618690 Avena lon
19	1218.6	47.6	2280	8	AY618691	AY618691 Avena lon
20	1218.6	47.6	2280	8	AY618692	AY618692 Avena pro
21	1217	47.5	2280	8	AY618693	AY618693 Avena str
22	1215.4	47.5	2280	8	AY618694	AY618694 Avena ven
23	1212.2	47.4	2280	8	AST311790	AJ311790 Avena str
24	1202.8	47.0	2579	8	AB025353	AB025353 Allium ma
25	1200.8	46.9	3057	8	AF216755	AF216755 Abies mag
26	1198.6	46.8	1983	8	AB025344	AB025344 Olea euro
27	1183.6	46.2	2558	6	AR407876	AR407876 Sequence
28	1176.6	46.0	2487	8	AB058508	AB058508 Costus sp
29	1086	42.4	2444	8	AB116238	AB116238 Cucurbita
30	1071	41.8	2578	8	AK066327	AK066327 Oryza sat
31	1051.8	41.1	2684	8	AB009031	AB009031 Panax gin
32	993.2	38.8	2347	8	AB033335	AB033335 Luffa cyl
33	980.4	38.3	2471	8	AB055511	AB055511 Betula pl
34	975.4	38.1	2546	8	AB025343	AB025343 Olea euro
35	967.6	37.8	2458	8	AK103080	AK103080 Oryza sat
36	963	37.6	2558	8	AK068026	AK068026 Oryza sat
37	960	37.5	2519	8	AB055512	AB055512 Betula pl
38	955.6	37.3	2493	8	AB025346	AB025346 Taraxacum
39	944.2	36.9	2478	6	AX382059	AX382059 Sequence
40	944	36.9	1882	6	AR407877	AR407877 Sequence
41	937.6	36.6	2729	8	AB014057	AB014057 Panax gin
42	929.4	36.3	2452	8	AB025345	AB025345 Taraxacum
43	916.4	35.8	2589	8	AB009030	AB009030 Panax gin
44	907	35.4	2475	8	AF478454	AF478454 Lotus jap
45	905.8	35.4	2671	8	AB037203	AB037203 Glycyrrhi

ALIGNMENTS

RESULT 1
AR407878
LOCUS AR407878 2560 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 5 from patent US 6630617.
ACCESSION AR407878
VERSION AR407878.1 GI:40157830
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 2560)
AUTHORS Famodu,O.O., Kinney,A.J. and Rafalski,J.A.
TITLE Enzymes involved in squalene metabolism
JOURNAL Patent: US 6630617-A 5 07-OCT-2003;
FEATURES Location/Qualifiers
source 1. .2560
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN

Query Match 100.0%; Score 2560; DB 6; Length 2560;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2560; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 TTGGCCTCTTGCCAGCAAAACAGAATGTGGAAGCTCAAGTTCGCCGAAGGAGGGAATCCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TTGGCCTCTTGCCAGCAAAACAGAATGTGGAAGCTCAAGTTCGCCGAAGGAGGGAATCCA 60

Qy         61 TGGCTTCGGACATTGAACAATCACGTTGGAAGACAGGTGTGGGAGTTCGATCCTAAGCTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGGCTTCGGACATTGAACAATCACGTTGGAAGACAGGTGTGGGAGTTCGATCCTAAGCTT 120

Qy        121 GGATCGCCGCAAGATCTCCTCGAGATTGAGAAAGCTCGCCAGAATTTTCACGATAACCGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATCGCCGCAAGATCTCCTCGAGATTGAGAAAGCTCGCCAGAATTTTCACGATAACCGC 180

Qy        181 TTTACCCACAAACACAGCGCTGATCTACTTATGCGGATGCAGTTCGCAAGAGAGAACCCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTTACCCACAAACACAGCGCTGATCTACTTATGCGGATGCAGTTCGCAAGAGAGAACCCA 240

Qy        241 ACACGTGAAGTCTTGCCCAAAGTCGGAGTTAAGGATATTGAGGATGTGACCCAAGAGATT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

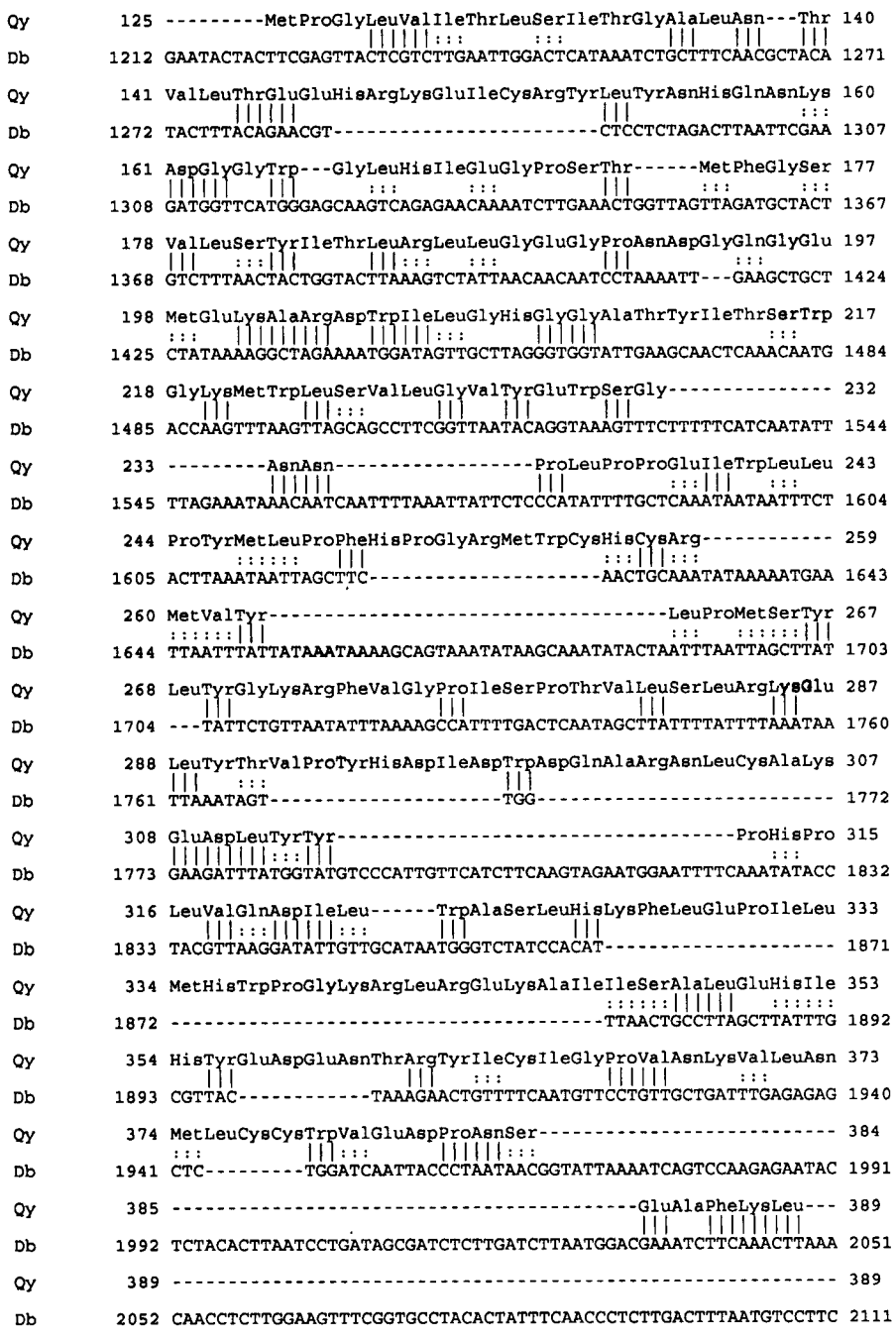
```
              :::|||||||||::::::                    |||   |||      :::
Db       3110 ACTATTAGCTCTTTTAGAAGTTTATAATTAAAATGAACAAATTAAGCATGCTGCAGAT-- 3167

Qy        699 oLeuHisArgAlaAlaAlaTyrLeuIleAsn---SerGlnLeuGluAspGlyAspPhePr 718
                  |||||||||   ||||||::::::      :::   :::|||      |||
Db       3168 -------AGAGCTGCTTAGTATTTATTAGATTAATTCAAAAGAGACGATAATACCTTCTA 3220

Qy        718 oGlnGlnGluIleMetGlyValPheAsnLysAsnCysMet---IleThrTyrAlaAlaTy 737
                      :::|||      ::::::      :::   :::   |||   :::|||
Db       3221 TGATCACTCCACAATAGGAACAGGTCACAGAGGATTACTCTATTTATAGTACCCCTCATA 3280

Qy        737 rArgAsnIleProIleTrpAlaLeuGlyGluTyr 749
              |         ||||||:::   ||||||      |||
Db       3281 TGCACAATCATTCCCATTAGTAGCTTTAAATAGATAC 3317


Search completed: November  5, 2005, 17:20:42
Job time : 601 secs
```


Qy 125 ---------MetProGlyLeuValIleThrLeuSerIleThrGlyAlaLeuAsn---Thr 140
Db 1212 GAATACTACTTCGAGTTACTCGTCTTGAATTGGACTCATAAATCTGCTTTCAACGCTACA 1271
Qy 141 ValLeuThrGluGluHisArgLysGluIleCysArgTyrLeuTyrAsnHisGlnAsnLys 160
Db 1272 TACTTTACAGAACGT----------CTCCTCTAGACTTAATTCGAA 1307
Qy 161 AspGlyGlyTrp---GlyLeuHisIleGluGlyProSerThr------MetPheGlySer 177
Db 1308 GATGGTTCATGGGAGCAAGTCAGAGAACAAAATCTTGAAACTGGTTAGTTAGATGCTACT 1367
Qy 178 ValLeuSerTyrIleThrLeuArgLeuLeuGlyGluGlyProAsnAspGlyGlnGlyGlu 197
Db 1368 GTCTTTAACTACTGGTACTTAAAGTCTATTAACAACAATCCTAAAATT---GAAGCTGCT 1424
Qy 198 MetGluLysAlaArgAspTrpIleLeuGlyHisGlyGlyAlaThrTyrIleThrSerTrp 217
Db 1425 CTATAAAAGGCTAGAAAATGGATAGTTGCTTAGGGTGGTATTGAAGCAACTCAAACAATG 1484
Qy 218 GlyLysMetTrpLeuSerValLeuGlyValTyrGluTrpSerGly---------- 232
Db 1485 ACCAAGTTTAAGTTAGCAGCCTTCGGTTAATACAGGTAAAGTTTCTTTTTCATCAATATT 1544
Qy 233 ---------AsnAsn----------ProLeuProProGluIleTrpLeuLeu 243
Db 1545 TTAGAAATAAACAATCAATTTTAAATTATTCTCCCATATTTTGCTCAAATAATAATTTCT 1604
Qy 244 ProTyrMetLeuProPheHisProGlyArgMetTrpCysHisCysArg---------- 259
Db 1605 ACTTAAATAATTAGCTTC----------AACTGCAAATATAAAAATGAA 1643
Qy 260 MetValTyr----------LeuProMetSerTyr 267
Db 1644 TTAATTTATTATAAATAAAAGCAGTAAATATAAGCAAATATACTAATTTAATTAGCTTAT 1703
Qy 268 LeuTyrGlyLysArgPheValGlyProIleSerProThrValLeuSerLeuArgLysGlu 287
Db 1704 ---TATTCTGTTAATATTTAAAAGCCATTTTGACTCAATAGCTTATTTTATTTTAAATAA 1760
Qy 288 LeuTyrThrValProTyrHisAspIleAspTrpAspGlnAlaArgAsnLeuCysAlaLys 307
Db 1761 TTAAATAGT----------TGG---------- 1772
Qy 308 GluAspLeuTyrTyr----------ProHisPro 315
Db 1773 GAAGATTTATGGTATGTCCCATTGTTCATCTTCAAGTAGAATGGAATTTTCAAATATACC 1832
Qy 316 LeuValGlnAspIleLeu------TrpAlaSerLeuHisLysPheLeuGluProIleLeu 333
Db 1833 TACGTTAAGGATATTGTTGCATAATGGGTCTATCCACAT---------- 1871
Qy 334 MetHisTrpProGlyLysArgLeuArgGluLysAlaIleIleSerAlaLeuGluHisIle 353
Db 1872 ----------TTAACTGCCTTAGCTTATTTG 1892
Qy 354 HisTyrGluAspGluAsnThrArgTyrIleCysIleGlyProValAsnLysValLeuAsn 373
Db 1893 CGTTAC----------TAAAGAACTGTTTTCAATGTTCCTGTTGCTGATTTGAGAGAG 1940
Qy 374 MetLeuCysCysTrpValGluAspProAsnSer---------- 384
Db 1941 CTC---------TGGATCAATTACCCTAATAACGGTATTAAAATCAGTCCAAGAGAATAC 1991
Qy 385 ----------GluAlaPheLysLeu--- 389
Db 1992 TCTACACTTAATCCTGATAGCGATCTCTTGATCTTAATGGACGAAATCTTCAAACTTAAA 2051
Qy 389 ---------- 389
Db 2052 CAACCTCTTGGAAGTTTCGGTGCCTACACTATTTCAACCCTCTTGACTTTAATGTCCTTC 2111
Qy 390 ----------HisLeuProArgIleTyr------AspTyrLeuTrpIleAla 401
Db 2112 AAAGACTTTTAGTCAAAGCAC---CCTCATCTATACCAAAACGAAATACAAAAGGCTTAC 2168
Qy 402 GluAspGlyMetLysMet----------GlnGlyTyrAsnGly 412
Db 2169 GAAGACGGATACTATTTCGTTGAGTTTAACTACTTTAACTTTAGAGAAGCTTATCACGGC 2228
Qy 413 Ser----------GlnLeuTrpAspThrAlaPheAlaValGlnAlaIleIleAlaSer 428
Db 2229 TCTTTGGATGATGGTAGATGGTGGGATACCATTCTTATTAGTTGGGCTATGCTTGAAAGT 2288
Qy 429 ----------AsnLeuIleGluGluPheGlyPro 436
Db 2289 GGCTAAGATAAAGAAAGAATCTTCCCTATCGTATAAAATATGGTCAAAGAA---GGTCTT 2345
Qy 437 ThrIleArgLysAlaHisThrTyrIleLysAsnSerGlnValLeuGluAspCysProGly 456
Db 2346 TAACCTAAAAAAGGTATAGGTTATGGATATGATTTCGAATATGCTCCTGACACTGATGAC 2405
Qy 457 Asp----------LeuAsnLysTrpTyrArgHisIle---------- 465
Db 2406 ACTGGATTACTTCTCGTTGTTATGAGTTACTACAAAGAAGCCTTCTAAAAGTAAATCCCT 2465
Qy 466 SerLysGlyAlaTrpProPheSer------ThrGlyAspHisGlyTrpProIleSerAsp 483
Db 2466 GAAACTATTGAATGGCTTTTCTCTATGCAAAATGACGATGGTGGCTATCCAGCTTTTGAC 2525
Qy 484 CysThrAlaGluGlyLeuLysAlaValLeuLeuLeuSerLysIleAlaProGluIleVal 503
Db 2526 ----------AAAGGTAATTTAATATTGATAATTTATTCCATTTCTTTATTT 2567
Qy 504 GlyGlu----------ProIleAspValLysArgLeuTyrAsp 514
Db 2568 AATAAAAATAAATCTTTTAATTATTTCAATTGAAAGATACATTTAAATAAAATTACAAAT 2627
Qy 515 ------SerValAsnValIleLeuSerLeu----------GlnAsn 524
Db 2628 GTACTTAAAATAAATATAATATTATTAACACTTCTACTTTATTTTAAAATAGGTAAAAAT 2687
Qy 525 GluAspGlyGlyPheAlaThrTyrGluLeuLysArgSerTyrAsnTrpLeuGluIleIle 544
Db 2688 GAAGAC----------AATTTATTGTTCAAGTTTGCCTTCAATATGGCTGGTATTGCT 2735
Qy 545 AsnProAlaGluThrPheGlyAspIleValIleAspTyrProTyrValGluCysThrSer 564
Db 2736 AACTCAGCTGAAATCTTC----------GATCCCTCATGTCCTGATATTACTGGT 2780
Qy 565 AlaAlaIleGlnAlaLeuAlaSerPheArgLysLeuTyrProGlyHisArgArgGluGlu 584
Db 2781 CACATCATGGAAGGATTGGGTGAGTTT----------GGATAT---------CAA 2816
Qy 585 IleGlnHisCys---IleAspLysAlaThrThrPheIleGluLysIleGlnAlaSerAsp 603
Db 2817 GCTAATCATCCTTAGATTTAAAATATGATTAAATATCAAAGAAAGACTTAGAACAAGTGG 2876
Qy 604 GlySerTrpTyrGlySerTrpGlyValCysPheThrTyrGlyAlaTrp-PheGlyValLy 623
Db 2877 GGATCTTGGTAAGCTAGATGGGGTGTAAATTACAT-TATGGCTGTTGGAGCTGTTGTACC 2935
Qy 623 sGlyLeuIleAlaAlaGlyArgSerPheSerAsnCysSerSerIleArgLysAlaCysGl 643
Db 2936 AGGTTTAGCAAGAGTCAATTACGACTTA---AATGAACAGTGGGTACAAAATAGTATAAA 2992
Qy 643 uPheLeuLeuSerLysGlnLeuProSerGlyGlyTrpGlyGluSerTyrLeuSerCysGl 663
Db 2993 TTATTTGCTTAATAAATAAAATAAAGATGGTGGCTTTGGTGAATGTGTCCTTTCTTATAA 3052
Qy 663 nAsnLysValTyrSerAsnLeuGluGlyAsnArgSerHisValValAsnThrGlyTrpAl 683
Db 3053 TGATCCTGAAAAGTGGAATGGTATAGGT---AAATCTACTGTCACTCAAACCTCATGGGG 3109
Qy 683 aMetLeuAlaLeuIleAspAla----------GlyGlnAlaLysArgAspSerGlnPr 699

```
Db      24315 TCTGTGGAGGCTGCTGTGTATCCAGAGCGCCTGCAGGGAGGTGTCTTTGGGGACTGTGGG 24374

Qy        323 ----------------------------AlaSerLeuHisLysPheLeuGluProIle 332

Db      24375 GACTGTGGGGACCCATGCCATGGGCAGTAGGCTGCTGTGTGTGCATGGTTGCCACCG--- 24431

Qy        333 LeuMetHisTrpProGlyLysArgLeuArgGluLysAlaIleIleSerAlaLeuGluHis 352

Db      24432 ------TACTGGTCTTGGGGGAGG-----------------ATCTCAGCCCTGGTCCAC 24467

Qy        353 IleHisTyrGluAspGluAsnThrArgTyrIleCysIleGlyProValAsnLysValLeu 372

Db      24468 CTC---------------------------------------------------------- 24470

Qy        373 AsnMetLeuCysCysTrpValGluAspProAsnSerGluAlaPheLysLeuHisLeuPro 392

Db      24471 --------------TGGGCACCT-------------------------CACATACCC 24488

Qy        393 ArgIleTyrAspTyrLeuTrpIleAlaGluAspGlyMetLysMetGlnGlyTyrAsn-Gl 412

Db      24489 --------------GCCTTCCTGGTCCCCTCCACATCACACATGGCTTTTTGGGGTGGG 24533

Qy        412 ySerGlnLeuTrpAspThrAlaPheAlaValGlnAlaIleIleAlaSerAsnLeuIleGl 432

Db      24534 GTCGCAGCTT---TCTGCTGTGTTCCCCTCATCTTCGCTCTCAGGTAGCACAGGTGTG-- 24588

Qy        432 uGluPheGlyProThrIleArgLysAlaHisThrTyrIleLysAsnSerGlnValLeuGl 452

Db      24588 ---------------------------------------------------------- 24588

Qy        452 uAspCysProGlyAspLeuAsnLysTrpTyrArgHisIleSerLysGlyAlaTrpProPh 472

Db      24589 ----TGTCCTGGA-----------------------------------------CCAGC 24602

Qy        472 eSerThrGlyAspHisGlyTrpProIleSerAspCysThrAlaGluGlyLeuLysAlaVa 492

Db      24603 CGGCGTTTGCTCTGGAGGTTGGTCAGG-GAGGCAGCGTCCGGGCCCGGGCTCACTGCAAC 24661

Qy        492 lLeuLeuLeuSerLysIleAlaProGluIleValGlyGluProIleAspValLysArgLe 512

Db      24662 ACTCTTGCTTGTTGTGGCTTTGCCTGAGCTGCAGAGC---------------------- 24698

Qy        512 uTyrAspSerValAsnValIleLeuSerLeuGlnAsnGluAspGlyGlyPheAlaThrTy 532

Db      24699 ----------------------CTGGGCAGCCAGGGTGAAACC----------------- 24719

Qy        532 rGluLeuLysArgSerTyrAsnTrpLeuGluIleIleAsnProAlaGluThrPheGlyAs 552

Db      24720 ------------------------CAACACTTGGTTCTTCCCTCCCTTTCCCA-GGGGA 24753

Qy        552 pIleValIleAspTyrProTyrValGluCysThrSerAlaAlaIleGlnAlaLeuAlaSe 572

Db      24754 CATCATGATTGACTACACCTATGTGGAGTGCACCTCAGCCGTGATGCAGGCGCTTAAGTA 24813

Qy        572 rPheArgLysLeuTyrProGlyHisArgArgGluGluIleGln 586

Db      24814 TTTCCACAAGCGTTTCCCGGAGCACAGGGCAGCGGAGATCCGG 24856
```

RESULT 14
US-09-248-796A-3810
; Sequence 3810, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/096,410
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 3810
; LENGTH: 255
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-3810

Alignment Scores:

Pred. No.:	2.19e-08	Length:	255
Score:	151.00	Matches:	30
Percent Similarity:	61.84%	Conservative:	17
Best Local Similarity:	39.47%	Mismatches:	27
Query Match:	3.64%	Indels:	2
DB:	4	Gaps:	1

US-09-419-679-6 (1-757) x US-09-248-796A-3810 (1-255)

```
Qy        677 ValValAsnThrGlyTrpAlaMetLeuAlaLeuIleAspAlaGlyGlnAlaLysArgAsp 696

Db         10 GTTGTTCGATCGGCTTGGGCTTTGATAGGGTTGATATTGGGAAATTATCCT------GAT 63

Qy        697 SerGlnProLeuHisArgAlaAlaAlaTyrLeuIleAsnSerGlnLeuGluAspGlyAsp 716

Db         64 GAAGAGCCAATCAAAAGAGGAATTCAATTTTTGATGAAACGACAATTACCGACTGGTGAG 123

Qy        717 PheProGlnGlnGluIleMetGlyValPheAsnLysAsnCysMetIleThrTyrAlaAla 736

Db        124 TGGAAGTATGAAGATATTGAAGGTGTATTTAATCACAGTTGTGCAATAGAATATCCTTCA 183

Qy        737 TyrArgAsnIlePheProIleTrpAlaLeuGlyGluTyrGlnSerGln 752

Db        184 TATAGGTTTTTATTTCCAATTAAGGCATTAGGTTTATATAAAAACAAG 231
```

RESULT 15
US-09-725-735A-13
; Sequence 13, Application US/09725735A
; Patent No. 6645751
; GENERAL INFORMATION:
; APPLICANT: RUSING, MATTHIAS
; APPLICANT: SCHWEINS, THOMAS
; APPLICANT: DRESLER, PETRA
; APPLICANT: STOCK, WOLFGANG
; APPLICANT: KIY, THOMAS
; TITLE OF INVENTION: NOVEL NUCLEIC ACID ISOLATED FROM TETRAHYMENA WHICH
; TITLE OF INVENTION: CODES FOR A TRITERPENOID CYCLASE, ITS PRODUCTION AND
; TITLE OF INVENTION: USE
; FILE REFERENCE: 25426.0001
; CURRENT APPLICATION NUMBER: US/09/725,735A
; CURRENT FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: DE 199 57 889.3
; PRIOR FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 5046
; TYPE: DNA
; ORGANISM: Tetrahymena thermophila
US-09-725-735A-13

Alignment Scores:

Pred. No.:	1.86e-05	Length:	5046
Score:	145.50	Matches:	172
Percent Similarity:	33.83%	Conservative:	103
Best Local Similarity:	21.16%	Mismatches:	276
Query Match:	3.51%	Indels:	263
DB:	4	Gaps:	47

US-09-419-679-6 (1-757) x US-09-725-735A-13 (1-5046)

```
Qy        108 GlnCysHisAspGlyHisTrp------ProGlyAspTyrGlyGlyProMetPheLeu--- 124

Db       1155 GAAGTACACAACAATACTTGGTACTATCCT---CCTTACTTAGGCGAAATGTTTATCAGT 1211
```

```
Db       1483 AGATGGGGTGTAAATTACAT-TATGGCTGTTGGAGCTGTTGTACCAGGTTTAGCAAGAGT 1541

Qy        628 aGlyArgSerPheSerAsnCysSerSerIleArgLysAlaCysGluPheLeuLeuSerLy 648

Db       1542 CAATTACGACTTA---AATGAACAGTGGGTACAAAATAGTATAAATTATTTGCTTAATAA 1598

Qy        648 sGlnLeuProSerGlyGlyTrpGlyGluSerTyrLeuSerCysGlnAsnLysValTyrSe 668

Db       1599 ATAAAATAAAGATGGTGGCTTTGGTGAATGTGTCCTTTCTTATAATGATCCTGAAAAGTG 1658

Qy        668 rAsnLeuGluGlyAsnArgSerHisValValAsnThrGlyTrpAlaMetLeuAlaLeuIl 688

Db       1659 GAATGGTATAGGT---AAATCTACTGTCACTCAAACCTCATGGGGACTATTAGCTCTTTT 1715

Qy        688 eAspAla------------GlyGlnAlaLysArgAspSerGlnProLeuHisArgAlaAl 704

Db       1716 AGAAGTTTATAATTAAAATGAACAAATTAAGCATGCTGCAGAT---------AGAGCTGC 1766

Qy        704 aAlaTyrLeuIleAsn---SerGlnLeuGluAspGlyAspPheProGlnGlnGluIleMe 723

Db       1767 TTAGTATTTATTAGATTAATTCAAAAGAGACGATAATACCTTCTATGATCACTCCACAAT 1826

Qy        723 tGlyValPheAsnLysAsnCysMet---IleThrTyrAlaAlaTyrArgAsnIlePhePr 742

Db       1827 AGGAACAGGTCACAGAGGATTACTCTATTTATAGTACCCCTCATATGCACAATCATTCCC 1886

Qy        742 oIleTrpAlaLeuGlyGluTyr 749

Db       1887 ATTAGTAGCTTTAAATAGATAC 1908
```

RESULT 12
US-09-313-294A-804
; Sequence 804, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath V.
; APPLICANT: Ito, Laura Y.
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
; FILE REFERENCE: PL-0017 US
; CURRENT APPLICATION NUMBER: US/09/313,294A
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 7600
; SOFTWARE: PERL Program
; SEQ ID NO 804
; LENGTH: 173
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6476212 700549863H1
US-09-313-294A-804

Alignment Scores:			
Pred. No.:	2.49e-10	Length:	173
Score:	165.00	Matches:	32
Percent Similarity:	83.93%	Conservative:	15
Best Local Similarity:	57.14%	Mismatches:	7
Query Match:	3.98%	Indels:	2
DB:	4	Gaps:	1

US-09-419-679-6 (1-757) x US-09-313-294A-804 (1-173)

```
Qy        489 LeuLysAlaValLeuLeuLeuSerLysIleAlaProGlu------IleValGlyGluPro 506

Db          5 CTTAAGGCTGTACTGTTGCTGTCAAAGATCTCAAACAAAAACAACCTTGTTGGGGATCCA 64

Qy        507 IleAspValLysArgLeuTyrAspSerValAsnValIleLeuSerLeuGlnAsnGluAsp 526

Db         65 ATAGAAAGAGAAAGGTTGCATGACGCTATTGATTGCCTTCTATCTTTTGCGAACAAAGAT 124

Qy        527 GlyGlyPheAlaThrTyrGluLeuLysArgSerTyrAsnTrpLeuGlu 542

Db        125 GGCACCTTTTCTACATATGAGTGCAAAAGAACTTATTCTTGGTTAGAG 172
```

RESULT 13
US-09-815-048-3
; Sequence 3, Application US/09815048
; Patent No. 6737261
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001180
; CURRENT APPLICATION NUMBER: US/09/815,048
; CURRENT FILING DATE: 2001-03-23
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 42450
; TYPE: DNA
; ORGANISM: Human
US-09-815-048-3

Alignment Scores:			
Pred. No.:	2.67e-05	Length:	42450
Score:	158.00	Matches:	109
Percent Similarity:	31.11%	Conservative:	45
Best Local Similarity:	22.02%	Mismatches:	150
Query Match:	3.81%	Indels:	195
DB:	4	Gaps:	20

US-09-419-679-6 (1-757) x US-09-815-048-3 (1-42450)

```
Qy        124 LeuMetProGlyLeuValIleThrLeuSerIle---------ThrGlyAlaLeuAsnThr 140

Db      23856 CTGTGCCCTGGGCGCCTGTGTGTGCTCAGCACATGGGCGAGTTCTAGGGTGCTCTCTGTG 23915

Qy        141 ValLeuThrGluGluHisArgLysGluIleCysArgTyrLeuTyrAsnHisGlnAsnLys 160

Db      23916 ATTTCAGCTGCTGAACATGAGAAATCCAGATGGAGGGTTCGCCACC--TATGAGACCAAG 23973

Qy        161 AspGlyGlyTrpGlyLeuHisIleGluGlyProSerThrMetPheGlySerValLeuSer 180

Db      23974 CGTGGGGGGCACTTGCTGGAGCTGCTGAACCCCTCGGAGGTCTTCGGT------------ 24021

Qy        181 TyrIleThrLeuArgLeuLeuGlyGluGlyProAsnAspGlyGlnGlyGluMetGluLys 200

Db      24021 ------------------------------------------------------------ 24021

Qy        201 AlaArgAspTrpIleLeuGlyHisGlyGlyAlaThrTyrIleThrSerTrpGlyLysMet 220

Db      24022 ------GAGTGGTCGGCCAG-CACTGCGGCGCG---CAAACCCGGGGCTGGCTAGCACTG 24071

Qy        221 TrpLeuSerValLeuGlyValTyrGluTrpSerGlyAsnAsnProLeuProPro------ 238

Db      24072 TGGTACACAAACCTGGGG-----------GCCAGCTTTTCCCCCTTGCCCGAGGCTGCA 24119

Qy        239 -----------------GluIleTrpLeuLeuProTyrMetLeuPro------------ 248

Db      24120 AGGGCCCAGGTTCACCGGCAGATCTGTCTGGAGCCCTCCCTCAGCCCAGGCTGTTCTGCG 24179

Qy        249 ---PheHisProGlyArgMetTrpCysHisCysArgMetValTyrLeuProMetSerTyr 267

Db      24180 CTCCTCCATCCCCCGGGGTGGCAGGATCCTTGT-------------------------- 24212

Qy        268 LeuTyrGlyLysArgPheValGlyProIleSerProThrValLeuSerLeuArgLysGlu 287

Db      24213 GTTGTGGATAGGAGGGCATCAGGTCAGACCTAGGGGACAGTGGAGGGT------------ 24260

Qy        288 LeuTyrThrValProTyrHisAspIleAspTrpAsp------GlnAlaArgAsnLeuCys 305

Db      24261 ------TCCAGTGAGATCCACAGCCTGGGCTGGTTCCTGCTCAGTCCACAGGGCTTGTGT 24314

Qy        306 AlaLysGluAspLeuTyrTyrProHisProLeuValGlnAspIleLeuTrp--------- 322
```

```
;  APPLICANT:  STOCK, WOLFGANG
;  APPLICANT:  KIY, THOMAS
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID ISOLATED FROM TETRAHYMENA WHICH
;  TITLE OF INVENTION:  CODES FOR A TRITERPENOID CYCLASE, ITS PRODUCTION AND
;  TITLE OF INVENTION:  USE
;  FILE REFERENCE: 25426.0001
;  CURRENT APPLICATION NUMBER: US/09/725,735A
;  CURRENT FILING DATE:  2000-11-30
;  PRIOR APPLICATION NUMBER: DE 199 57 889.3
;  PRIOR FILING DATE: 1999-12-01
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
;   LENGTH: 2072
;   TYPE: DNA
;   ORGANISM: Tetrahymena thermophila
US-09-725-735A-11

Alignment Scores:
Pred. No.:                2.58e-12      Length:          2072
Score:                    198.00        Matches:         160
Percent Similarity:       35.16%        Conservative:    96
Best Local Similarity:    21.98%        Mismatches:      247
Query Match:              4.77%         Indels:          226
DB:                       4             Gaps:            41

US-09-419-679-6 (1-757) x US-09-725-735A-11 (1-2072)

Qy        108 GlnCysHisAspGlyHisTrp------ProGlyAspTyrGlyGlyProMetPheLeu--- 124
Db        145 GAAGTACACAACAATACTTGGTACTATCCT---CCTTACTTAGGCGAAATGTTTATCAGT 201

Qy        125 ---------MetProGlyLeuValIleThrLeuSerIleThrGlyAlaLeuAsn---Thr 140
Db        202 GAATACTACTTCGAGTTACTCGTCTTGAATTGGACTCATAAATCTGCTTTCAACGCTACA 261

Qy        141 ValLeuThrGluGluHisArgLysGluIleCysArgTyrLeuTyrAsnHisGlnAsnLys 160
Db        262 TACTTTACAGAACGT----------------------CTCCTCTAGACTTAATTCGAA 297

Qy        161 AspGlyGlyTrp---GlyLeuHisIleGluGlyProSerThr------MetPheGlySer 177
Db        298 GATGGTTCATGGGAGCAAGTCAGAGAACAAAATCTTGAAACTGGTTAGTTAGATGCTACT 357

Qy        178 ValLeuSerTyrIleThrLeuArgLeuLeuGlyGluGlyProAsnAspGlyGlnGlyGlu 197
Db        358 GTCTTTAACTACTGGTACTTAAAGTCTATTAACAACAATCCTAAAATT---GAAGCTGCT 414

Qy        198 MetGluLysAlaArgAspTrpIleLeuGlyHisGlyGlyAlaThrTyrIleThrSerTrp 217
Db        415 CTATAAAAGGCTAGAAAATGGATAGTTGCTTAGGGTGGTATTGAAGCAACTCAAACAATG 474

Qy        218 GlyLysMetTrpLeuSerValLeuGlyValTyrGluTrpSerGlyAsnAsnProLeuPro 237
Db        475 ACCAAGTTTAAGTTAGCAGCCTTCGGTTAATACAGTTGGGAA----------------- 516

Qy        238 ProGluIleTrpLeuLeuProTyrMetLeuProPheHisProGlyArgMetTrpCysHis 257
Db        517 ---GATTTATGGTATGTCCCATTGTTCATC---TTCAAGTAGAATGGAATTTTC------ 564

Qy        258 CysArgMetValTyrLeuProMetSerTyrLeuTyrGlyLysArgPheValGly------ 275
Db        565 ----------------------AAATATACCTACGTTAAGGATATTGTTGCATAATGG 600

Qy        276 ------ProIleSerProThrValLeuSerLeuArg------LysGluLeuTyrThrVal 291
Db        601 GTCTATCCACATTTAACTGCCTTAGCTTATTTGCGTTACTAAAGAACTGTTTTCAATGTT 660

Qy        292 ProTyrHisAspIleAspTrpAspGlnAlaArgAsnLeuCysAlaLysGluAspLeuTyr 311
Db        661 CCTGTTGCTGATTTG----------------------------AGAGAG--------- 681

Qy        312 TyrProHisProLeuValGlnAspIleLeuTrpAlaSerLeuHisLysPheLeuGluPro 331
```

```
Db        682 -------------------------CTCTGG-------------------------- 687

Qy        332 IleLeuMetHisTrpProGlyLysArgLeuArgGluLysAlaIleIleSerAlaLeuGlu 351
Db        688 ------ATCAATTACCCTAAGAACGGTATTAAA-----------ATCAGTCCAAGAGAA 729

Qy        352 HisIleHisTyrGluAspGluAsnThrArgTyrIleCysIleGlyProValAsnLysVal 371
Db        730 ----------------------------TACTCTACACTTAATCCTGATAGCGATCTC 759

Qy        372 LeuAsnMetLeuCysCysTrpValGluAspProAsnSerGluAlaPheLysLeu------ 389
Db        760 TTGATCTTAATG----------------------GACGAAATCTTCAAACTTAAACAA 795

Qy        389 ------------------------------------------------------------ 389
Db        796 CCTCTTGGAAGTTTCGGTGCCTACACTATTTCAACCCTCTTGACTTTAATGTCCTTCAAA 855

Qy        390 --------------HisLeuProArgIleTyr------AspTyrLeuTrpIleAlaGlu 402
Db        856 GACTTTTAGTCAAAGCAC---CCTCATCTATACCAAAACGAAATACAAAAGGCTTACGAA 912

Qy        403 AspGlyMetLysMet-------------------------GlnGlyTyrAsnGlySer 413
Db        913 GACGGATACTATTTCGTTGAGTTTAACTACTTTAACTTTAGAGAAGCTTATCACGGCTCT 972

Qy        414 -----------GlnLeuTrpAspThrAlaPheAlaValGlnAlaIleIleAlaSerAsn 429
Db        973 TTGGATGATGGTAGATGGTGGGATACC----------------ATTCTTATTAGTTGG 1014

Qy        430 LeuIleGluGluPheGlyProThrIleArgLysAlaHisThrTyrIleLysAsnSerGln 449
Db       1015 GCTATGCTTGAAAGTGGCTAAGATAAAGAAAGAATCTTCCCTATCGTATAAAATATG--- 1071

Qy        450 ValLeuGluAspCysProGlyAspLeuAsnLysTrpTyrArgHisIleSerLysGlyAla 469
Db       1072 ------------------------------------------GTCAAAGAAGGTCTT 1086

Qy        470 TrpProPheSerThrGlyAspHisGlyTrp--------------ProIleSerAspCys 484
Db       1087 TAACCTAAAAAAGGTATAGGTTATGGATATGATTTCGAATATGCTCCTGACACTGATGAC 1146

Qy        485 ThrAlaGluGlyLeuLysAlaValLeuLeu----------------LeuSerLysIle 498
Db       1147 ACT------GGATTACTTCTCGTTGTTATGAGTTACTACAAAGAAGCCTTCTAAAAGTAA 1200

Qy        499 AlaProGluIleValGlyGluProIleAspValLysArgLeuTyrAspSerValAsnVal 518
Db       1201 ATCCCTGAA-------------------------------------ACTATTGAATGG 1221

Qy        519 IleLeuSerLeuGlnAsnGluAspGlyGlyPheAlaThrTyrGlu--------------- 533
Db       1222 CTTTTCTCTATGCAAAATGACGATGGTGGCTATCCAGCTTTTGACAAAGGTAAAAATGAA 1281

Qy        534 -----------LeuLysArgSerTyrAsnTrpLeuGluIleIleAsnProAlaGluThr 549
Db       1282 GACAATTTATTGTTCAAGTTTGCCTTCAATATGGCTGGTATTGCTAACTCAGCTGAAATC 1341

Qy        550 PheGlyAspIleValIleAspTyrProTyrValGluCysThrSerAlaAlaIleGlnAla 569
Db       1342 TTC--------------GATCCCTCATGTCCTGATATTACTGGTCACATCATGGAAGGA 1386

Qy        570 LeuAlaSerPheArgLysLeuTyrProGlyHisArgArgGluGluIleGlnHisCys--- 588
Db       1387 TTGGGTGAGTTT--------------GGATAT---------CAAGCTAATCATCCTTAG 1422

Qy        589 IleAspLysAlaThrThrPheIleGluLysIleGlnAlaSerAspGlySerTrpTyrGly 608
Db       1423 ATTTAAAATATGATTAAATATCAAAGAAAGACTTAGAACAAGTGGGGATCTTGGTAAGCT 1482

Qy        609 SerTrpGlyValCysPheThrTyrGlyAlaTrp-PheGlyValLysGlyLeuIleAlaAl 628
```

```
Qy        560 ValGluCysThrSerAlaAlaIleGlnAlaLeuAlaSerPheArgLysLeuTyrProGly 579
Db     535265 GAAGACGTTACCGCACGGTGCGTCTCGATGTTGGCCCAGGTCGGCGAATTC--------- 535215

Qy        580 HisArgArgGluGluIleGlnHisCysIleAspLysAlaThrThrPheIleGluLysIle 599
Db     535214 ---------ACCCAAAGGAGCAAGGCAGTCGCCGAGGGGATTGCATATTTGCGCCGCACC 535164

Qy        600 GlnAlaSerAspGlySerTrpTyrGlySerTrpGlyValCysPheThrTyrGlyAlaTrp 619
Db     535163 CAACACGCGGAAGGATCATGGTACGGCCGCTGGGGCCTGAACTACATCTACGGTACCTGG 535104

Qy        620 PheGlyValLysGlyLeuIleAlaAlaGlyArgSerPheSerAsnCysSerSerIleArg 639
Db     535103 TCGGTGTTGTGCGCGCTGAATGCCGCAGGGATCGATCACCAAGAC---CCCATGATAAGG 535047

Qy        640 LysAlaCysGluPheLeuLeuSerLysGlnLeuProSerGlyGlyTrpGlyGluSerTyr 659
Db     535046 AAGGCGGTGGAATGGCTAGTATCCATCCAGAGTTGGGACGGCGGTTGGGGCGAGGATGCA 534987

Qy        660 LeuSerCysGlnAsnLysValTyrSerAsnLeuGluGlyAsnArgSerHisValValAsn 679
Db     534986 ATCAGCTACCGACTCGAC---TATAGCGGATATGAGCAGGCGCCTTCCACGTCCTCCCAA 534930

Qy        680 ThrGlyTrpAlaMetLeuAlaLeuIleAspAlaGlyGlnAlaLysArgAspSerGlnPro 699
Db     534929 ACGGCGTGGGCCTTGCTTGGACTGATGGCGGCCGGCGAGGTGGAGCATCCGGCC------ 534876

Qy        700 LeuHisArgAlaAlaAlaTyrLeuIleAsnSerGlnLeuGluAspGlyAspPheProGln 719
Db     534875 GTCGCACGCGGGGTGAACTACCTAAAAAACGCACAAACCGAGAACGGTCTGTGGGATGAG 534816

Qy        720 GlnGluIleMetGlyVal---PheAsnLysAsnCysMetIleThrTyrAlaAlaTyrArg 738
Db     534815 CAGCGCTACACCGCCACGGGTTTTCCGCGGGTGTTTTATTTGCGATATCACGGCTACTCC 534756

Qy        739 AsnIlePheProIleTrpAlaLeuGlyGluTyrGlnSer 751
Db     534755 AAGTTCTTTCCGCTCTGGGCATTAGCACGGTACCGGAAC 534717

RESULT 10
US-08-998-416-229/c
; Sequence 229, Application US/08998416
; Patent No. 6239264
;  GENERAL INFORMATION:
;    APPLICANT:  Philippsen, Peter
;    APPLICANT:  Pohlmann, Rainer
;    APPLICANT:  Steiner, Sabine
;    APPLICANT:  Mohr, Christine
;    APPLICANT:  Wendland, Jurgen
;    APPLICANT:  Knechtle, Philipp
;    APPLICANT:  Rebischung, Corinne
;    TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  1152
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6239264artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  No. 6239264th Carolina
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/998,416
;      FILING DATE:  24-DEC-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 0016/97
;      FILING DATE:  31-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  PF/5-30306/A/CGC1976
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  229:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  539 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  PAG1211RP
US-08-998-416-229

Alignment Scores:
Pred. No.:              4.11e-30      Length:          539
Score:                  340.50        Matches:         77
Percent Similarity:     58.38%        Conservative:    24
Best Local Similarity:  44.51%        Mismatches:      58
Query Match:            8.21%         Indels:          15
DB:                     3             Gaps:            5

US-09-419-679-6 (1-757) x US-08-998-416-229 (1-539)

Qy        112 GlyHisTrpProGlyAspTyrGlyGlyProMetPheLeuMetProGlyLeuValIleThr 131
Db        504 GGGCANTTTCCTGTTGTAACAGGGGGCCNGATGTNATGATAT---GGGTATGTGGTGGCG 448

Qy        132 LeuSerIleThrGlyAlaLeuAsnThrValLeuThrGluGluHisArgLysGluIleCys 151
Db        447 ATGCATGTGGGCGGG---------GTGGTATTCCT-GAGCCTCAGCGGCGGGAGCTAGTG 398

Qy        152 ArgTyrLeuTyrAsnHisGlnAsnLys---AspGlyGlyTrpGlyLeuHisIleGluGly 170
Db        397 CGCTAACGTGGTGAACACGGCACACCTGTCGACGGCGGGTGGGGATTGCATTCGGAGGAC 338

Qy        171 ProSerThrMet-PheGlySerValLeuSerTyrIleThrLeuArgLeuLeuGlyGluGl 190
Db        337 AAGTCGACTGTCGATGGGCACCGCGCTGAACTACGTTCTGCTGCGGCTCTTAGGGATG-- 280

Qy        190 yProAsnAspGlyGlnGlyGlu---------MetGluLysAlaArgAspTrpIleLeuGl 207
Db        279 ------------GAGCGCGAACACATCGTGATGCAGCGGGCGCGGAAGACCCTGCTGCG 233

Qy        207 yHisGlyGlyAlaThrTyrIleThrSerTrpGlyLysMetTrpLeuSerValLeuGlyVa 227
Db        232 CATGGGGGGTGCTATTGGCTCGCCGTACTGGGGTAAAGTGTGGCTTTCCGTGCTCAATTT 173

Qy        227 lTyrGluTrpSerGlyAsnAsnProLeuProProGluIleTrpLeuLeuProTyrMetLe 247
Db        172 ATACAAGTGGGAGGGCGTGAACCCAGCGCCACCAGAGCTATGGCTGCTGCCGTACCAGTT 113

Qy        247 uProPheHisProGlyArgMetTrpCysHisCysArgMetValTyrLeuProMetSerTy 267
Db        112 TCCTTATCACCCGGCTCGCTGGTGGGTACATACGCGTGCTATCTTTTTGCCGATGAGCTA 53

Qy        267 rLeuTyrGlyLysArgPheValGlyProIleSerPro 279
Db         52 CCTGGCAGGGGTCCGCTACCAGTGTGCTCTGACGCCA 16

RESULT 11
US-09-725-735A-11
; Sequence 11, Application US/09725735A
; Patent No. 6645751
; GENERAL INFORMATION:
;  APPLICANT: RUSING, MATTHIAS
;  APPLICANT:  SCHWEINS, THOMAS
;  APPLICANT:  DRESLER, PETRA
```

```
Db         61 ATACTTGAGCCGTATAGAGTTGAGATGATACGTTATATTGTCAACACTGCACACCCAGTC 120

Qy        161 AspGlyGlyTrpGlyLeuHisIleGluGlyProSerThrMetPheGlySerValLeuSer 180
Db        121 GATGGTGGTTGGGGACTTCATTCTGTTGATAAATATACTTGTTTTGGGACAACCATGAAT 180

Qy        181 TyrIleThrLeuArgLeuLeuGlyGluGlyProAsnAspGlyGlnGlyGluMetGlu--- 199
Db        181 TATGTATGTCTCCGGTTATTAGGA-------------------------ATGGAAAAG 213

Qy        200 ----------------LysAlaArgAspTrpIleLeuGlyHisGlyGlyAlaThrTyr 213
Db        214 GATCATCCAGTTTTGGTTAAGGCGAGAAAAACATTACATCGTTTGGGTGGTGCCATTAAG 273

Qy        214 IleThrSerTrpGlyLysMetTrpLeuSerValLeuGlyValTyrGluTrpSerGlyAsn 233
Db        274 AATCCACATTGGGGTAAGGCTTGGCTATCTATTTTGAATTTATATGAATGGGAGGGTGTG 333

Qy        234 AsnProLeuProProGluIleTrpLeuLeuProTyrMetLeuProPheHisProGlyArg 253
Db        334 AACCCAGCTCCACCAGAACTTTGGAGATTACCGTACTGGTTACCAATTCATCCAGCGAAA 393

Qy        254 MetTrpCysHisCysArgMetValTyrLeuProMetSerTyrLeuTyrGlyLysArgPhe 273
Db        394 TGGTGGGTACATACTAGGGCTATCTATTTGCCATTGGGATATACGTCTGCAAACAGAGTT 453

Qy        274 ValGlyProIleSerProThrValLeuSerLeuArgLysGluLeuTyrThr--------- 290
Db        454 CAATGTGAGCTTGATCCACTTTTAAAAGAGATCAGAAATGAAATTTACGTTCCAAGTCAA 513

Qy        291 ValProTyrHisAspIleAspTrpAspGlnAlaArgAsnLeuCysAlaLysGluAspLeu 310
Db        514 TTACCTTATGAGTCAATCAAATTTGGTAACCAGAGAAATAATGTTTGTGGTGTCGATTTA 573

Qy        311 TyrTyrProHisProLeuValGlnAspIleLeuTrpAlaSerLeuHisLysPheLeuGlu 330
Db        574 TATTACCCACATACAAAGATTCTTGATTTTGCAAATTCTATATTGAGTAAATGG---GAA 630

Qy        331 ProIle------------LeuMetHisTrpProGlyLysArgLeu 341
Db        631 GCTGTTAGACCTAAGTGGTTATTGAATTGGGTTAACAAGAAAGTT 675
```

RESULT 9
US-09-214-808-1/c
; Sequence 1, Application US/09214808A
; Patent No. 6475793
; GENERAL INFORMATION:
; APPLICANT: Rosenthal, Andre
; APPLICANT: Freiberg, Christoph
; APPLICANT: Perret, Xavier Philippe
; APPLICANT: Broughton, William John
; TITLE OF INVENTION: Genomic Sequence of Rhizobium SP. NGR 234 Symbiotic
; Patent No. 6475793
; TITLE OF INVENTION: Plasmid
; FILE REFERENCE: CARP0068
; CURRENT APPLICATION NUMBER: US/09/214,808A
; CURRENT FILING DATE: 1999-06-22
; PRIOR APPLICATION NUMBER: PCT/IB97/00950
; PRIOR FILING DATE: 1997-07-10
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 536165
; TYPE: DNA
; ORGANISM: Rhizobium
US-09-214-808-1

Alignment Scores:

Pred. No.:	3.01e-27	Length:	536165
Score:	361.00	Matches:	145
Percent Similarity:	42.03%	Conservative:	79
Best Local Similarity:	27.20%	Mismatches:	235
Query Match:	8.70%	Indels:	74
DB:	4	Gaps:	22

US-09-419-679-6 (1-757) x US-09-214-808-1 (1-536165)

```
Qy        242 LeuLeuProTyrMetLeuProPheHisProGlyArgMet------TrpCysHisCysArg 259
Db     536162 CTACTGCCAATCTGGTCGCCGTTCCACCTCTACAAGATCTCCTACTGGGCGCGCACCACT 536103

Qy        260 MetValTyrLeuProMetSerTyrLeuTyrGlyLysArgPheValGlyProIleSerPro 279
Db     536102 ATTGTGCCGCTGATGGTTCTTGCAGTGCTGAAGCCGCGTGCGAAGAATCCTAAGGGCGTC 536043

Qy        280 ThrValLeuSerLeuArgLysGluLeuTyrThrValProTyrHisAspIleAspTrpAsp 299
Db     536042 GGCATC-----------GAAGAACTGTTCCTTCAGGATACCAAGAGCGTAGGCATGAAC 535995

Qy        300 GlnAlaArgAsnLeuCysAlaLysGluAspLeuTyrTyrProHisProLeuValGlnAsp 319
Db     535994 CCGAAGGCG----------------------------CCGCAC---------CAGAGC 535974

Qy        320 IleLeuTrpAlaSerLeuHisLysPheLeuGluProIleLeuMet-----------His 335
Db     535973 TGGGGCTGGTTCTTGCTGTTCCGCGGCATCGACGGCATCCTGCGGGTTATTGAGCCGCAT 535914

Qy        336 TrpProGlyLysArgLeuArgGluLysAlaIleIleSerAlaLeu-----------Glu 351
Db     535913 CTTCCG---AAGAAACTTCGTGAGCGCGCGATCGCGAGTGCGCTCGCCTTCACTGAAGAG 535857

Qy        352 HisIleHisTyrGluAspGluAsnThrArgTyrIleCysIleGlyProValAsnLysVal 371
Db     535856 CGGCTCAACGGTGAAGAT---GGCATGGGCGCGATCTATCCGTCGATGGCCAATATCGTG 535800

Qy        372 LeuAsnMetLeuCysCysTrpValGluAspProAsnSerGluAlaPheLysLeuHisLeu 391
Db     535799 ATGATGTATGACGCGCTCGGTAAAGACGAT----------------------CATTTT 535764

Qy        392 Pro--------------------ArgIleTyrAspTyrLeuTrpIleAlaGluAspGly 404
Db     535763 CCGCCGCGCGCGATAGCACGTCGCGCTATTGACAAGCTCTTGGTGATCGGCGAGGAAGAG 535704

Qy        405 MetLysMetGlnGlyTyrAsnGlySerGlnLeuTrpAspThrAlaPhe---------Ala 421
Db     535703 GCCTATTGCCAGCCCTGCCTG---TCGCCGGTCTGGGACACCGCGCTGACGTGCCACGCA 535647

Qy        422 ValGlnAlaIleIleAlaSerAsnLeuIleGluGluPheGlyProThrIleArgLysAla 441
Db     535646 CTTCAGGAGGTGGGCGGCGCTAACGCCGTGGCGAAA-----------GCGAAGCAGGGC 535599

Qy        442 HisThrTyrIleLysAsnSerGlnValLeuGluAspCysProGlyAspLeuAsnLysTrp 461
Db     535598 CTGGACTGGCTGAAGCCGCGGCAGGTACTG---GACGTGAAGGGCGACTGGGCAGTGAAG 535542

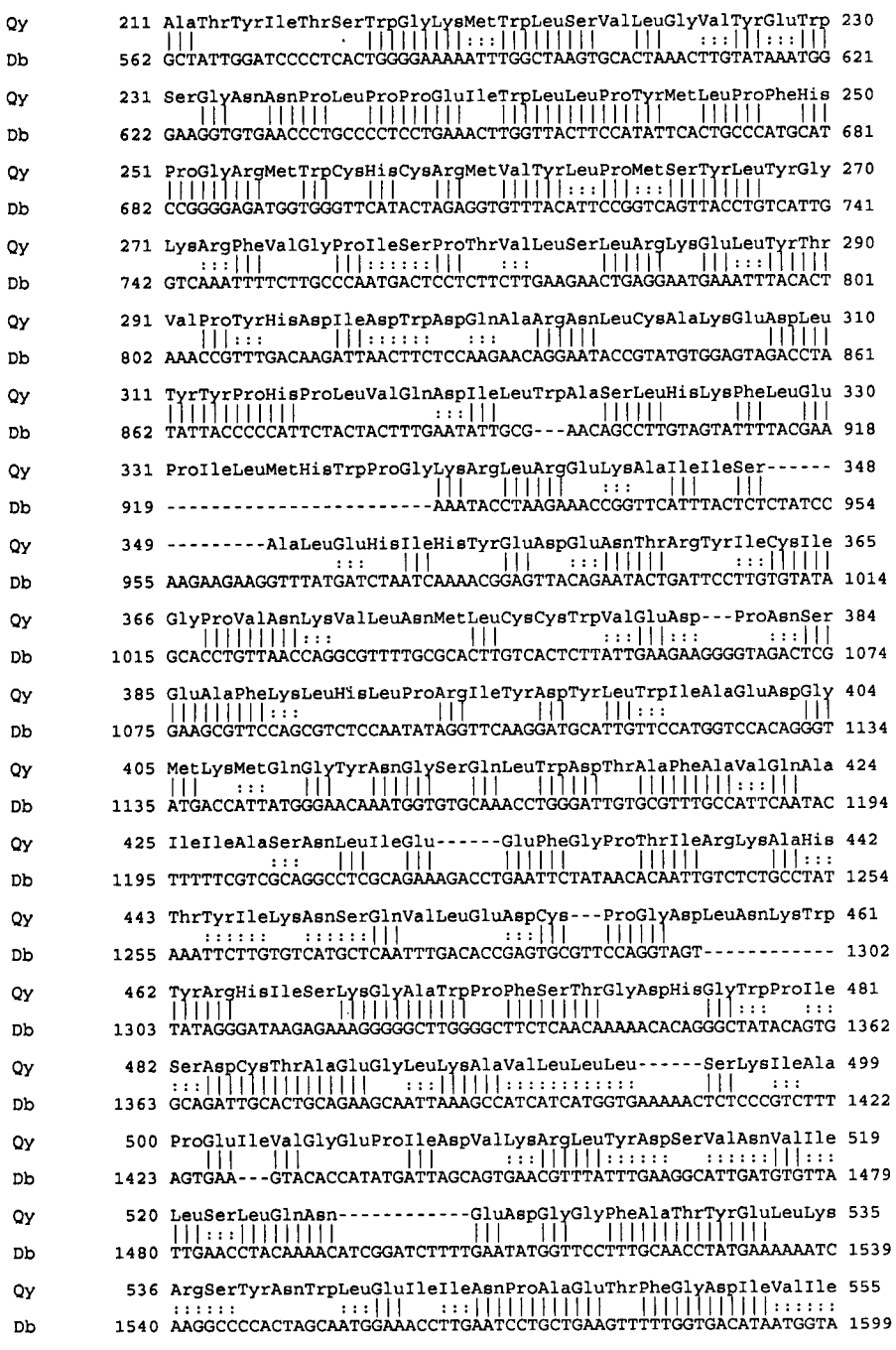
Qy        462 TyrArgHisIleSerLysGlyAlaTrpProPheSerThrGlyAspHisGlyTrpProIle 481
Db     535541 GCGCCCAACATCAGGCCCGGCGGCTGGCCGTTTCAATACAACAACGCGCACTATCCCGAT 535482

Qy        482 SerAspCysThrAlaGluGlyLeuLysAlaValLeuLeuLeuSerLysIleAlaProGlu 501
Db     535481 CTCGATGACACTGCA-----------GTGGTGGTCATGGCCATGGACCGCGCGCAGCGG 535434

Qy        502 IleValGlyGluProIleAspValLysArgLeuTyrAspSerValAsnValIleLeuSer 521
Db     535433 CACGCCGGCAGCAAGGAATACGCCACCGCGATCGCACGTGGCCGGGAATGGATCGAAGGC 535374

Qy        522 LeuGlnAsnGluAspGlyGlyPheAlaThrTyrGluLeuLysArgSerTyrAsnTrpLeu 541
Db     535373 ATGCAAAGCCGGGATGGCGGCTGGGCCGCCTTCGACGTC-----------AACAATCTC 535326

Qy        542 GluIleIleAsnProAlaGluThrPheGlyAsp------IleValIleAspTyrProTyr 559
Db     535325 GAATACTACCTGAATAACCTCCCATTCGCGGACCATGGCGCGCTGCTCGACCCGCCGACC 535266
```



RESULT 8
US-09-248-796A-3811
; Sequence 3811, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICA
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 3811
; LENGTH: 747
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-3811

Alignment Scores:

Pred. No.:	2.82e-37	Length:	747
Score:	403.50	Matches:	84
Percent Similarity:	53.02%	Conservative:	30
Best Local Similarity:	39.07%	Mismatches:	76
Query Match:	9.73%	Indels:	25
DB:	4	Gaps:	6

US-09-419-679-6 (1-757) x US-09-248-796A-3811 (1-747)



```
US-09-419-679-6 (1-757) x US-09-419-679-7 (1-1300)

Qy        417 AspThrAlaPheAlaValGlnAlaIleIleAlaSerAsnLeuIleGluGluPheGlyPro 436
Db          8 GACACAGCTTTTGCAGTTCAAGCTATTGCGGCCACTGACCTCATTGAAGAGTTTGCTCCC 67

Qy        437 ThrIleArgLysAlaHisThrTyrIleLysAsnSerGlnValLeuGluAspCysProGly 456
Db         68 ACTCTTAAGCTGGCACATGATTTTATTAAGAACTCTCAGGTTGTTGATGACTGCCCTGGA 127

Qy        457 AspLeuAsnLysTrpTyrArgHisIleSerLysGlyAlaTrpProPheSerThrGlyAsp 476
Db        128 GATCTGAGTTACTGGTACCGTCACATATCTAAAGGTGCATGGCCCTTTTCTACAGCTGAT 187

Qy        477 HisGlyTrpProIleSerAspCysThrAlaGluGlyLeuLysAlaValLeuLeuLeuSer 496
Db        188 CATGGTTGGCCTATATCAGATTGCACTGCAGAAGGACTAAAGGCCTCATTATTGCTATCA 247

Qy        497 LysIleAlaProGluIleValGlyGluProIleAspValLysArgLeuTyrAspSerVal 516
Db        248 AAGATTTCTCCAGAAATTGTGGGCGAATCGGTGGAAGTTAACAGACTATATGATGCTGTC 307

Qy        517 AsnValIleLeuSerLeuGlnAsnGluAspGlyGlyPheAlaThrTyrGluLeuLysArg 536
Db        308 AATTGTTTGATGTCTTGGATGAATGAAAATGGTGGCTTCGCAACATATGAACTCCAAAGG 367

Qy        537 SerTyrAsnTrpLeuGluIleIleAsnProAlaGluThrPheGlyAspIleValIleAsp 556
Db        368 TTTTATGCCTGGCTTGAGCTTATCAACCCTGCCGAGACATTCGGAGATATTGTGATTGAT 427

Qy        557 TyrProTyrValGluCysThrSerAlaAlaIleGlnAlaLeuAlaSerPheArgLysLeu 576
Db        428 TACCCGTATGTGGAATGTACCTCAGCCGCAATTCAGGCCCTGACATCATTTAAAAAGCTC 487

Qy        577 TyrProGlyHisArgArgGluGluIleGlnHisCysIleAspLysAlaThrThrPheIle 596
Db        488 TATCCTGGGCACCGCAGGAAAGATGTAGATAACTGTATCAACAAAGCTGCTAGTTACATT 547

Qy        597 GluLysIleGlnAlaSerAspGlySerTrpTyrGlySerTrpGlyValCysPheThrTyr 616
Db        548 GAGAGCATCCAAAGAAAAGATGGTTCATGGTATGGCTCTTGGGCTGTGTGCTTCACCTAT 607

Qy        617 GlyAlaTrpPheGlyValLysGlyLeuIleAlaAlaGlyArgSerPheSerAsnCysSer 636
Db        608 GGCACATGGTTCGGAGTGAAGGGGCTACTAGCTGCAGGTAGAACCTTCAAGAGCAGTCCT 667

Qy        637 SerIleArgLysAlaCysGluPheLeuLeuSerLysGlnLeuProSerGlyGlyTrpGly 656
Db        668 GCAATCAGAAAGGCATGTGACTTTCTGATGTCAAAAGAGCTTCCTTTCGGTGGCTGGGGA 727

Qy        657 GluSerTyrLeuSerCysGlnAsnLysValTyrSerAsnLeuGluGlyAsnArgSerHis 676
Db        728 GAAAGCTATCTGTCATCTCAAGATCAGGTTTACACCAATCTTGAAGGGAAGCATACTCAT 787

Qy        677 ValValAsnThrGlyTrpAlaMetLeuAlaLeuIleAspAlaGlyGlnAlaLysArgAsp 696
Db        788 GCTGTCAACACTGGCTGGGCCATGCTGACTCTAATTGACGCAGGACAGGCTGAGAGAGAC 847

Qy        697 SerGlnProLeuHisArgAlaAlaAlaTyrLeuIleAsnSerGlnLeuGluAspGlyAsp 716
Db        848 CCGACGCCTCTGCATCGAGCAGCGAAGGTTTTGATAAACTTACAATCAGAGGATGGGGAA 907

Qy        717 PheProGlnGlnGluIleMetGlyValPheAsnLysAsnCysMetIleThrTyrAlaAla 736
Db        908 TTTCCTCAGCAAGAGATCATGGGAGTCTTCAACAAGAACTGCATGATCAGCTACTCCCAG 967

Qy        737 TyrArgAsnIleProIleTrpAlaLeuGlyGluTyrGlnSerGlnValLeu 754
Db        968 TATCGGAACATCTTCCCTATCTGGGCGCTTGGCGAGTACCGCTGCCGGGTGCTG 1021

RESULT 7
US-09-614-221A-308

; Sequence 308, Application US/09614221A
; Patent No. 6723837
; GENERAL INFORMATION:
;  APPLICANT: Karunanandaa, Balasulojini
;  APPLICANT:  Yu, Jaehyuk
;  APPLICANT:  Kishore, Ganesh M.
;  TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED
;  TITLE OF INVENTION:  WITH STEROL SYNTHESIS AND METABOLISM
;  FILE REFERENCE: 16516.075
;  CURRENT APPLICATION NUMBER: US/09/614,221A
;  CURRENT FILING DATE:  2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/142,981
;  PRIOR FILING DATE: 1999-07-12
;  NUMBER OF SEQ ID NOS: 626
; SEQ ID NO 308
;   LENGTH: 2196
;   TYPE: DNA
;   ORGANISM: Saccharomyces cerevisiae
US-09-614-221A-308

Alignment Scores:
Pred. No.:              4.05e-139       Length:         2196
Score:                  1276.00         Matches:        303
Percent Similarity:     53.88%          Conservative:   114
Best Local Similarity:  39.15%          Mismatches:     263
Query Match:            30.76%          Indels:         94
DB:                     4               Gaps:           23

US-09-419-679-6 (1-757) x US-09-614-221A-308 (1-2196)

Qy          1 MetTrpLysLeuLysPheAlaGluGlyGlyAsnProTrpLeuArgThrLeuAsnAsnHis 20
Db         52 CTTTGGAGACTGAGA----------------------------------ACTGATGAG 75

Qy         21 ValGlyArgGlnValTrpGluPhe----------------------AspProLysLeu 32
Db         76 CTAGGCCGAGAAAGCTGGGAATATTTAACCCCTCAGCAAGCCGCAAACGACCCACCATCC 135

Qy         33 GlySerProGlnAspLeuLeuGluIleGluLysAlaArgGlnAsnPheHisAspAsnArg 52
Db        136 ACTTTCACGCAGTGGCTTCTTCAAGATCCCAAATTTCCTCAACCT---CATCCAGAAAGA 192

Qy         53 PheThrHisLysHisSerAlaAspLeuLeuMetArgMetGlnPheAlaArgGluAsnPro 72
Db        193 ------AATAAGCATTCACCAGATTTT---------------------------------- 213

Qy         73 ThrArgGluValLeuProLysValGlyValLysAspIleGluAspValThrGlnGluIle 92
Db        214 ----------------------------------------------------------TCA 216

Qy         93 ValThrLysThrLeuArgArgAlaValSerPheHisSerThrLeuGlnCysHisAsp--- 111
Db        217 GCCTTCGATGCGTGTCATAATGGTGCATCTTTTTTCAAACTGCTTCAAGAGCCTGACTCA 276

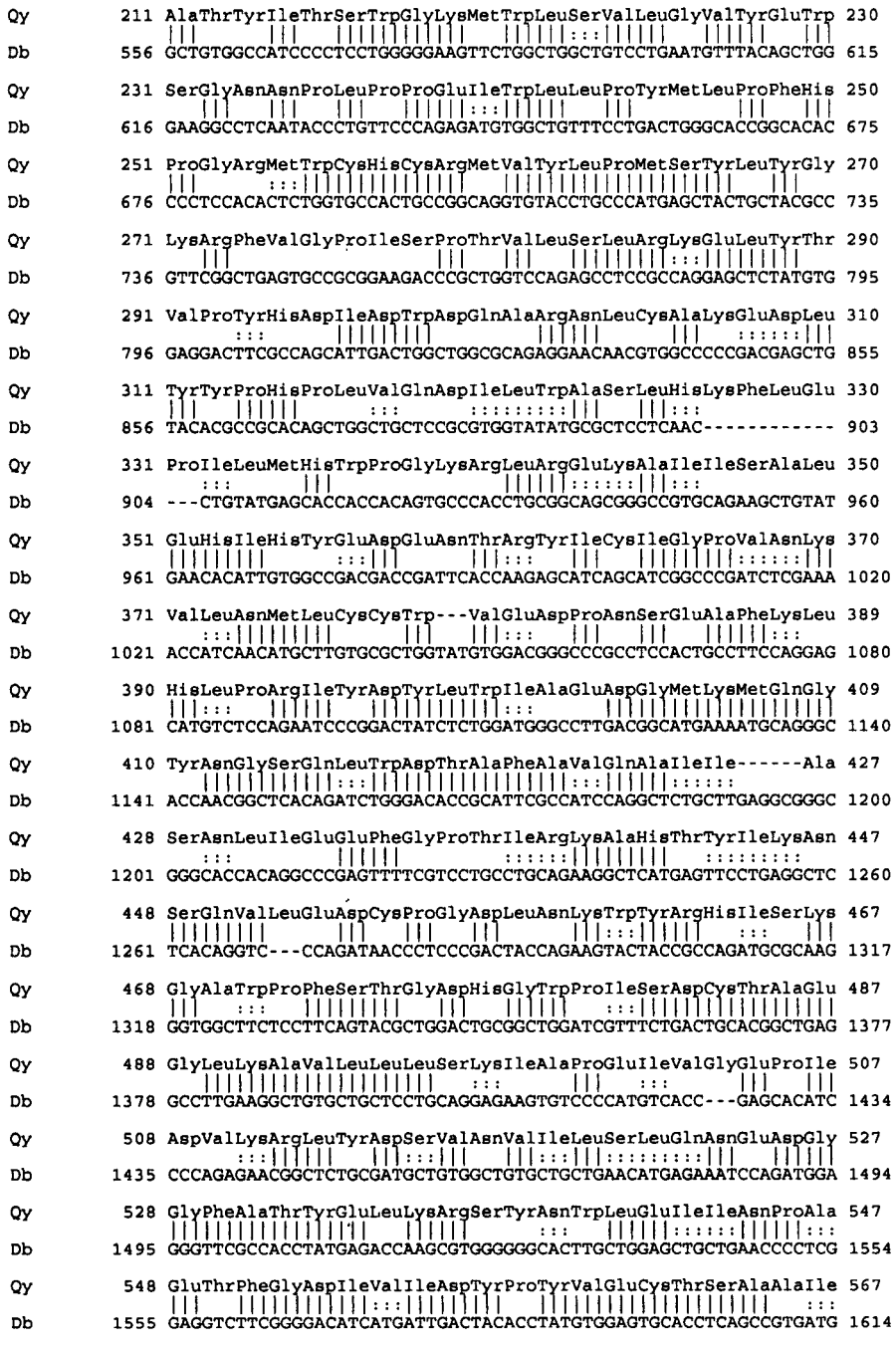
Qy        112 GlyHisTrpProGlyAspTyrGlyGlyProMetPheLeuMetProGlyLeuValIleThr 131
Db        277 GGTATTTTTCCGTGTCAATATAAAGGACCCATGTTCATGACAATCGGTTACGTAGCCGTA 336

Qy        132 LeuSerIleThrGlyAlaLeuAsnThrValLeuThrGluGluHisArgLysGluIleCys 151
Db        337 AACTATATCGCCGGT---------ATTGAAATTCCTGAGCATGAGAGAATAGAATTAATT 387

Qy        152 ArgTyrLeuTyrAsnHisGlnAsnLys---AspGlyGlyTrpGlyLeuHisIleGluGly 170
Db        388 AGATACATCGTCAATACAGCACATCCGGTTGATGGTGGCTGGGGTCTACATTCTGTTGAC 447

Qy        171 ProSerThrMetPheGlySerValLeuSerTyrIleThrLeuArgLeuLeuGlyGluGly 190
Db        448 AAATCCACCGTGTTTGGTACAGTATTGAACTATGTAATCTTACGTTTATTGGGTCTA--- 504

Qy        191 ProAsnAspGlyGlnGlyGluMetGluLysAlaArgAspTrpIleLeuGlyHisGlyGly 210
Db        505 CCCAAGGAC---CACCCGGTTTGCGCCAAGGCAAGAAGCACATTGTTAAGGTTAGGCGGT 561
```



RESULT 6
US-09-419-679-7
; Sequence 7, Application US/09419679
; Patent No. 6630617
; GENERAL INFORMATION:
; APPLICANT: Falco, S. Carl
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Hitz, William D.
; APPLICANT: Kinney, Anthony J.
; APPLICANT: Rafalski, Antoni
; APPLICANT: McGonigle, Brian
; APPLICANT: Lohman, Karin
; TITLE OF INVENTION: Enzymes Involved In Squalene Metabolizm
; FILE REFERENCE: BB1113 US NA
; CURRENT APPLICATION NUMBER: US/09/419,679
; CURRENT FILING DATE: 1999-10-15
; EARLIER APPLICATION NUMBER: 60/105,405
; EARLIER FILING DATE: 1998-10-23
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 7
; LENGTH: 1300
; TYPE: DNA
; ORGANISM: Triticum aestivum
US-09-419-679-7

Alignment Scores:

Pred. No.:	4.71e-159	Length:	1300
Score:	1442.00	Matches:	257
Percent Similarity:	88.46%	Conservative:	42
Best Local Similarity:	76.04%	Mismatches:	39
Query Match:	34.76%	Indels:	0
DB:	4	Gaps:	0

```
Qy        390 HisLeuProArgIleTyrAspTyrLeuTrpIleAlaGluAspGlyMetLysMetGlnGly 409
Db       1239 CATGTCTCCAGAATCCCGGACTATCTCTGGATGGGCCTTGACGGCATGAAAATGCAGGGC 1298

Qy        410 TyrAsnGlySerGlnLeuTrpAspThrAlaPheAlaValGlnAlaIleIle------Ala 427
Db       1299 ACCAACGGCTCACAGATCTGGGACACCGCATTCGCCATCCAGGCTCTGCTTGAGGCGGGC 1358

Qy        428 SerAsnLeuIleGluGluPheGlyProThrIleArgLysAlaHisThrTyrIleLysAsn 447
Db       1359 GGGCACCACAGGCCCGAGTTTTCGTCCTGCCTGCAGAAGGCTCATGAGTTCCTGAGGCTC 1418

Qy        448 SerGlnValLeuGluAspCysProGlyAspLeuAsnLysTrpTyrArgHisIleSerLys 467
Db       1419 TCACAGGTC---CCAGATAACCCTCCCGACTACCAGAAGTACTACCGCCAGATGCGCAAG 1475

Qy        468 GlyAlaTrpProPheSerThrGlyAspHisGlyTrpProIleSerAspCysThrAlaGlu 487
Db       1476 GGTGGCTTCTCCTTCAGTACGCTGGACTGCGGCTGGATCGTTTCTGACTGCACGGCTGAG 1535

Qy        488 GlyLeuLysAlaValLeuLeuLeuSerLysIleAlaProGluIleValGlyGluProIle 507
Db       1536 GCCTTGAAGGCTGTGCTGCTCCTGCAGGAGAAGTGTCCCCATGTCACC---GAGCACATC 1592

Qy        508 AspValLysArgLeuTyrAspSerValAsnValIleLeuSerLeuGlnAsnGluAspGly 527
Db       1593 CCCAGAGAACGGCTCTGCGATGCTGTGGCTGTGCTGCTGAACATGAGAAATCCAGATGGA 1652

Qy        528 GlyPheAlaThrTyrGluLeuLysArgSerTyrAsnTrpLeuGluIleIleAsnProAla 547
Db       1653 GGGTTCGCCACCTATGAGACCAAGCGTGGGGGGCACTTGCTGGAGCTGCTGAACCCCTCG 1712

Qy        548 GluThrPheGlyAspIleValIleAspTyrProTyrValGluCysThrSerAlaAlaIle 567
Db       1713 GAGGTCTTCGGGGACATCATGATTGACTACACCTATGTGGAGTGCACCTCAGCCGTGATG 1772

Qy        568 GlnAlaLeuAlaSerPheArgLysLeuTyrProGlyHisArgArgGluGluIleGlnHis 587
Db       1773 CAGGCGCTTAAGTATTTCCACAAGCGTTTCCCGGAGCACAGGGCAGCGGAGATCCGGGAG 1832

Qy        588 CysIleAspLysAlaThrThrPheIleGluLysIleGlnAlaSerAspGlySerTrpTyr 607
Db       1833 ACCCTCACGCAGGGCTTAGAGTTCTGTCGGCGGCAGCAGAGGGCCGATGGCTCCTGGGAA 1892

Qy        608 GlySerTrpGlyValCysPheThrTyrGlyAlaTrpPheGlyValLysGlyLeuIleAla 627
Db       1893 GGCTCCTGGGGAGTTTGCTTCACCTACGGCACCTGGTTTGGCCTGGAGGCCTTCGCCTGT 1952

Qy        628 AlaGlyArgSerPheSerAsn---------CysSerSerIleArgLysAlaCysGluPhe 644
Db       1953 ATGGGGCAGACCTACCGAGATGGGACTGCCTGTGCAGAGGTCTCCCGGGCCTGTGACTTC 2012

Qy        645 LeuLeuSerLysGlnLeuProSerGlyGlyTrpGlyGluSerTyrLeuSerCysGlnAsn 664
Db       2013 CTGCTGTCCCGGCAGATGGCAGACGGAGGCTGGGGGGAGGACTTTGAGTCCTGCGAGGAG 2072

Qy        665 LysValTyrSerAsnLeuGluGlyAsnArgSerHisValValAsnThrGlyTrpAlaMet 684
Db       2073 CGGCGTTAT------TTGCAGAGTGCCCAGTCCCAGATCCATAACACATGCTGGGCCATG 2126

Qy        685 LeuAlaLeuIleAspAlaGlyGlnAlaLysArgAspSerGlnProLeuHisArgAlaAla 704
Db       2127 ATGGGGCTGATGGCCGTTCGGCATCCTGACATCGAGGCCCAG------GAGAGAGGAGTC 2180

Qy        705 AlaTyrLeuIleAsnSerGlnLeuGluAspGlyAspPheProGlnGlnGluIleMetGly 724
Db       2181 CGGTGTCTACTTGAGAAACAGCTCCCCAATGGCGACTGGCCGCAGGAAAACATTGCTGGG 2240

Qy        725 ValPheAsnLysAsnCysMetIleThrTyrAlaAlaTyrArgAsnIlePheProIleTrp 744
Db       2241 GTCTTCAACAAGTCCTGTGCCATCTCCTACACGAGCTACAGGAACATCTTCCCCATCTGG 2300

Qy        745 AlaLeuGlyGluTyrGlnSerGlnValLeuGlnSerArg 757
Db       2301 GCCCTCGGCCGCTTC---TCCCAGCTGTACCCTGAGAGA 2336
```

RESULT 5
US-09-815-048-1
; Sequence 1, Application US/09815048
; Patent No. 6737261
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001180
; CURRENT APPLICATION NUMBER: US/09/815,048
; CURRENT FILING DATE: 2001-03-23
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 4239
; TYPE: DNA
; ORGANISM: Human
US-09-815-048-1

Alignment Scores:

Pred. No.:	2.18e-176	Length:	4239
Score:	1597.00	Matches:	329
Percent Similarity:	59.10%	Conservative:	116
Best Local Similarity:	43.69%	Mismatches:	232
Query Match:	38.50%	Indels:	76
DB:	4	Gaps:	16

US-09-419-679-6 (1-757) x US-09-815-048-1 (1-4239)

```
Qy         17 LeuAsnAsnHisValGlyArgGlnValTrpGluPhe------AspProLysLeuGlySer 34
Db        112 CTCAACTGCGAGAGGGGCCGGCAGACGTGGACCTACCTGCAGGACGAGCGCGCCGGCCGC 171

Qy         35 ProGlnAspLeuLeuGluIleGlu-------------LysAlaArgGlnAsnPheHisAsp 50
Db        172 GAGCAGACCGGCCTGGAAGCCTACGCCCTGGGGCTGGACACCAAGAATTACTTTAAGGAC 231

Qy         51 AsnArgPheThrHisLysHisSerAlaAspLeuLeuMetArgMetGlnPheAlaArgGlu 70
Db        231 ------------------------------------------------------------ 231

Qy         71 AsnProThrArgGluValLeuProLysValGlyValLysAspIleGluAspValThrGln 90
Db        232 ------------------TTGCCCAAAGCC---------------------------- 243

Qy         91 GluIleValThrLysThrLeuArgArgAlaValSerPheHisSerThrLeuGlnCysHis 110
Db        244 CACACCGCCTTTGAGGGGGCTCTGAACGGGATGACATTTTACGTGGGGCTGCAGGCTGAG 303

Qy        111 AspGlyHisTrpProGlyAspTyrGlyGlyProMetPheLeuMetProGlyLeuValIle 130
Db        304 GATGGGCACTGGACGGGTGATTATGGTGGCCCACTTTTCCTCCTGCCAGGCCTCCTGATC 363

Qy        131 ThrLeuSerIleThrGlyAlaLeuAsnThrValLeuThrGluGluHisArgLysGluIle 150
Db        364 ACTTGCCACGTGGCA---------CGCATCCCTCTGCCAGCCGGATACAGAGAAGAGATT 414

Qy        151 CysArgTyrLeuTyrAsnHisGlnAsnLysAspGlyGlyTrpGlyLeuHisIleGluGly 170
Db        415 GTGCGGTACCTG------------------------------CGGCACATTGAGGAT 441

Qy        171 ProSerThrMetPheGlySerValLeuSerTyrIleThrLeuArgLeuGlyGluGly 190
Db        442 AAGTCCACCGTGTTTGGGACTGCGCTCAACTATGTGTCTCTCAGAATTCTGGGTGTTGGG 501

Qy        191 ProAsnAspGlyGlnGlyGluMetGluLysAlaArgAspTrpIleLeuGlyHisGlyGly 210
Db        502 CCTGACGAT------CCTGACCTGGTACGAGCCCGGAACATTCTTCACAAGAAAGGTGGT 555
```

```
Db       1035 TGCACTTCAGCAGCAATTCAGGCCCTGACAGCATTTAAAAAGCTCTACCCTGGACACCGC 1094

Qy        582 ArgGluGluIleGlnHisCysIleAspLysAlaThrThrPheIleGluLysIleGlnAla 601

Db       1095 AAGAGTGAAATAGACAACTGTATAAGCAAAGCTGCTAGCTTTATTGAGGGTATTCAAAAA 1154

Qy        602 SerAspGlySerTrpTyrGlySerTrpGlyValCysPheThrTyrGlyAlaTrpPheGly 621

Db       1155 AGCGATGGTTCATGGTATGGTTCTTGGGCTGTTTGTTTTACCTATGGCACATGGTTTGGT 1214

Qy        622 ValLysGlyLeuIleAlaAlaGlyArgSerPheSerAsnCysSerSerIleArgLysAla 641

Db       1215 GTAAAGGGATTAGTTGCTGCTGGTAGGACATTCAAAAACAGTCCTGCAATCAGAAAGGCA 1274

Qy        642 CysGluPheLeuLeuSerLysGlnLeuProSerGlyGlyTrpGlyGluSerTyrLeuSer 661

Db       1275 TGTGACTTTTTGTTGTCAAAAGAGCTTCCTTCTGGAGGCTGGGGAGAAAGCTATTTGTCA 1334

Qy        662 CysGlnAsnLysValTyrSerAsnLeuGluGlyAsnArgSerHisValValAsnThrGly 681

Db       1335 TCCCAAGATCAGGTTTATACCAATCTCGAAGGGAAGCGACCTCATGCGGTGAACACTGGT 1394

Qy        682 TrpAlaMetLeuAlaLeuIleAspAlaGlyGlnAlaLysArgAspSerGlnProLeuHis 701

Db       1395 TGGGCCATGCTAGCCCTAATCGATGCAGGGCAGGCTGAGAGAGATCCAATTCCTTTGCAT 1454

Qy        702 ArgAlaAlaAlaTyrLeuIleAsnSerGlnLeuGluAspGlyAspPheProGlnGlnGlu 721

Db       1455 CGAGCAGCGAAGGTTTTGATCAACTTACAATCGGAAGATGGTGAATTTCCCCAGCAAGAG 1514

Qy        722 IleMetGlyValPheAsnLysAsnCysMetIleThrTyrAlaAlaTyrArgAsnIlePhe 741

Db       1515 ATCATTGGAGTCTTCAACAAAAACTGCATGATCAGCTACTCCGAGTATAGAAACATCTTC 1574

Qy        742 ProIleTrpAlaLeuGlyGluTyrGlnSerGlnValLeu 754

Db       1575 CCTATTTGGGCCCTTGGGGAGTACCGTCGCCGCGTCTTG 1613
```

RESULT 4
US-09-919-039-112
; Sequence 112, Application US/09919039
; Patent No. 6727066
; GENERAL INFORMATION:
; APPLICANT: Kaser, Matthew R.
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
; FILE REFERENCE: PA-0035 US
; CURRENT APPLICATION NUMBER: US/09/919,039
; CURRENT FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 60/222,113
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 401
; SOFTWARE: PERL Program
; SEQ ID NO 112
; LENGTH: 4346
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6727066 225080.16
US-09-919-039-112

Alignment Scores:

Pred. No.:	5.11e-183	Length:	4346
Score:	1653.50	Matches:	337
Percent Similarity:	60.16%	Conservative:	116
Best Local Similarity:	44.75%	Mismatches:	235
Query Match:	39.86%	Indels:	65
DB:	4	Gaps:	15

US-09-419-679-6 (1-757) x US-09-919-039-112 (1-4346)

```
Qy         17 LeuAsnAsnHisValGlyArgGlnValTrpGluPhe------AspProLysLeuGlySer 34

Db        237 CTCAACTGCGAGAGGGGCCGGCAGACGTGGACCTACCTGCAGGACGAGCGCGCCGGCCGC 296

Qy         35 ProGlnAspLeuLeuGluIleGlu-----------LysAlaArgGlnAsnPheHisAsp 50

Db        297 GAGCAGACCGGCCTGGAAGCCTACGCCCTGGGGCTGGACACCAAGAATTACTTTAAGGAC 356

Qy         51 AsnArgPheThrHisLysHisSerAlaAspLeuLeuMetArgMetGlnPheAlaArgGlu 70

Db        356 ------------------------------------------------------------ 356

Qy         71 AsnProThrArgGluValLeuProLysValGlyValLysAspIleGluAspValThrGln 90

Db        357 -----------------TTGCCCAAAGCC---------------------------- 368

Qy         91 GluIleValThrLysThrLeuArgArgAlaValSerPheHisSerThrLeuGlnCysHis 110

Db        369 CACACCGCCTTTGAGGGGGCTCTGAACGGGATGACATTTTACGTGGGGCTGCAGGCTGAG 428

Qy        111 AspGlyHisTrpProGlyAspTyrGlyGlyProMetPheLeuMetProGlyLeuValIle 130

Db        429 GATGGGCACTGGACGGGTGATTATGGTGGCCCACTTTTCCTCCTGCCAGGCCTCCTGATC 488

Qy        131 ThrLeuSerIleThrGlyAlaLeuAsnThrValLeuThrGluGluHisArgLysGluIle 150

Db        489 ACTTGCCACGTGGCA---------CGCATCCCTCTGCCAGCCGGATACAGAGAAGAGATT 539

Qy        151 CysArgTyrLeuTyrAsnHisGlnAsnLysAspGlyGlyTrpGlyLeuHisIleGluGly 170

Db        540 GTGCGGTACCTGCGGTCAGTGCAGCTCCCTGACGGTGGCTGGGGCCTGCACATTGAGGAT 599

Qy        171 ProSerThrMetPheGlySerValLeuSerTyrIleThrLeuArgLeuLeuGlyGluGly 190

Db        600 AAGTCCACCGTGTTTGGGACTGCGCTCAACTATGTGTCTCTCAGAATTCTGGGTGTTGGG 659

Qy        191 ProAsnAspGlyGlnGlyGluMetGluLysAlaArgAspTrpIleLeuGlyHisGlyGly 210

Db        660 CCTGACGAT------CCTGACCTGGTACGAGCCCGGAACATTCTTCACAAGAAAGGTGGT 713

Qy        211 AlaThrTyrIleThrSerTrpGlyLysMetTrpLeuSerValLeuGlyValTyrGluTrp 230

Db        714 GCTGTGGCCATCCCCTCCTGGGGGAAGTTCTGGCTGGCTGTCCTGAATGTTTACAGCTGG 773

Qy        231 SerGlyAsnAsnProLeuProProGluIleTrpLeuLeuProTyrMetLeuProPheHis 250

Db        774 GAAGGCCTCAATACCCTGTTCCCAGAGATGTGGCTGTTTCCTGACTGGGCACCGGCACAC 833

Qy        251 ProGlyArgMetTrpCysHisCysArgMetValTyrLeuProMetSerTyrLeuTyrGly 270

Db        834 CCCTCCACACTCTGGTGCCACTGCCGGCAGGTGTACCTGCCCATGAGCTACTGCTACGCC 893

Qy        271 LysArgPheValGlyProIleSerProThrValLeuSerLeuArgLysGluLeuTyrThr 290

Db        894 GTTCGGCTGAGTGCCGCGGAAGACCCGCTGGTCCAGAGCCTCCGCCAGGAGCTCTATGTG 953

Qy        291 ValProTyrHisAspIleAspTrpAspGlnAlaArgAsnLeuCysAlaLysGluAspLeu 310

Db        954 GAGGACTTCGCCAGCATTGACTGGCTGGCGCAGAGGAACAACGTGGCCCCCGACGAGCTG 1013

Qy        311 TyrTyrProHisProLeuValGlnAspIleLeuTrpAlaSerLeuHisLysPheLeuGlu 330

Db       1014 TACACGCCCCACAGCTGGCTGCTCCGCGTGGTATATGCGCTCCTCAAC------------ 1061

Qy        331 ProIleLeuMetHisTrpProGlyLysArgLeuArgGluLysAlaIleIleSerAlaLeu 350

Db       1062 ---CTGTATGAGCACCACCACAGTGCCCACCTGCGGCAGCGGGCCGTGCAGAAGCTGTAT 1118

Qy        351 GluHisIleHisTyrGluAspGluAsnThrArgTyrIleCysIleGlyProValAsnLys 370

Db       1119 GAACACATTGTGGCCGACGACCGATTCACCAAGAGCATCAGCATCGGCCCGATCTCGAAA 1178

Qy        371 ValLeuAsnMetLeuCysCysTrp---ValGluAspProAsnSerGluAlaPheLysLeu 389

Db       1179 ACCATCAACATGCTTGTGCGCTGGTATGTGGACGGGCCCGCCTCCACTGCCTTCCAGGAG 1238
```

```
Qy        574 ArgLysLeuTyrProGlyHisArgArgGluGluIleGlnHisCysIleAspLysAlaThr 593
              :::||||||||||||||||||||||||||:::|||:::   :::||||||   |||||||:::
Db       1563 AAAAAACTATACCCTGGGCACCGCAGGAAAGAGGTGGATAACTGTATCAGCAAAGCTTCC 1622

Qy        594 ThrPheIleGluLysIleGlnAlaSerAspGlySerTrpTyrGlySerTrpGlyValCys 613
              |||||||||   ||||||   ||||||||||||||||||||||||   ||||||
Db       1623 AATTTCATCGAGAGTATTCAGAAAAGCGATGGTTCATGGTATGGCTCTTGGGCCGTCTGT 1682

Qy        614 PheThrTyrGlyAlaTrpPheGlyValLysGlyLeuIleAlaAlaGlyArgSerPheSer 633
              ||||||||||||   ||||||||||||||||||||||||||||||:::|||
Db       1683 TTCACATACGGCACTTGGTTTGGTGTGAAGGGACTAATTGCTGCTGGTAGAACATTTGAG 1742

Qy        634 AsnCysSerSerIleArgLysAlaCysGluPheLeuLeuSerLysGlnLeuProSerGly 653
              |||      :::|||||||||||||||:::|||||||||||||:::||||||||||||
Db       1743 AACAGTCCTGCAATTAGAAAGGCATGCGACTTTCTGTTGTCAAAAGAACTTCCTTCCGGT 1802

Qy        654 GlyTrpGlyGluSerTyrLeuSerCysGlnAsnLysValTyrSerAsnLeuGluGlyAsn 673
              |||||||||||||||||||||||||   |||:::::||||||:::||||||:::||||||
Db       1803 GGTTGGGGAGAAAGCTATTTGTCATCTCAAGACCAGGTTTACACCAATCTCAAAGGCAAC 1862

Qy        674 ArgSerHisValValAsnThrGlyTrpAlaMetLeuAlaLeuIleAspAlaGlyGlnAla 693
              |||   |||   |||||||||   ||||||||||||||||||||||||||||||||||||
Db       1863 CGGCCTCATGCGGTGAACACTAGTTGGGCCATGCTGGCGCTGATTGATGCGGGCCAGGCC 1922

Qy        694 LysArgAspSerGlnProLeuHisArgAlaAlaAlaTyrLeuIleAsnSerGlnLeuGlu 713
              :::||||||      ||||||||||||||||      |||||||||   |||   |||
Db       1923 GAGAGAGATCCAACGCCTCTACACCGAGCAGCAAAGGTTTTGATCAACTTACAATCAGAG 1982

Qy        714 AspGlyAspPheProGlnGlnGluIleMetGlyValPheAsnLysAsnCysMetIleThr 733
              ||||||:::||||||||||||||||||:::||||||||||||||||||||||||||:::
Db       1983 GACGGAGAATTTCCTCAGCAAGAGATCATAGGAGTGTTCAACAAGAACTGCATGATAAGC 2042

Qy        734 TyrAlaAlaTyrArgAsnIlePheProIleTrpAlaLeuGlyGluTyrGlnSerGlnVal 753
              |||:::   ||||||||||||||||||||||||||||||||||||:::   :::|||
Db       2043 TACTCCCAGTACAGGAACATCTTCCCGATTTGGGCTCTGGGTGAGTACCGGTGTCGAGTC 2102

Qy        754 Leu 754
              |||
Db       2103 TTG 2105

RESULT 3
US-09-419-679-3
; Sequence 3, Application US/09419679
; Patent No. 6630617
; GENERAL INFORMATION:
;  APPLICANT: Falco, S. Carl
;  APPLICANT: Famodu, Omolayo O.
;  APPLICANT: Hitz, William D.
;  APPLICANT: Kinney, Anthony J.
;  APPLICANT: Rafalski, Antoni
;  APPLICANT: McGonigle, Brian
;  APPLICANT: Lohman, Karin
;  TITLE OF INVENTION: Enzymes Involved In Squalene Metabolizm
;  FILE REFERENCE: BB1113 US NA
;  CURRENT APPLICATION NUMBER: US/09/419,679
;  CURRENT FILING DATE: 1999-10-15
;  EARLIER APPLICATION NUMBER: 60/105,405
;  EARLIER FILING DATE: 1998-10-23
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 3
;   LENGTH: 1882
;   TYPE: DNA
;   ORGANISM: Oryza sativa
US-09-419-679-3

Alignment Scores:
Pred. No.:              2.21e-275       Length:         1882
Score:                  2433.00         Matches:        419
Percent Similarity:     90.43%          Conservative:   63
Best Local Similarity:  78.61%          Mismatches:     51
Query Match:            58.65%          Indels:         0
DB:                     4               Gaps:           0

US-09-419-679-6 (1-757) x US-09-419-679-3 (1-1882)

Qy        222 LeuSerValLeuGlyValTyrGluTrpSerGlyAsnAsnProLeuProProGluIleTrp 241
              ||||||||||||||||||||:::::||||||||||||||||||:::|||||||||||||||
Db         15 CTTTCGGTGCTTGGTGTATTTGACTGGTCTGGCAACAACCCAGTGCCACCAGAAATATGG 74

Qy        242 LeuLeuProTyrMetLeuProPheHisProGlyArgMetTrpCysHisCysArgMetVal 261
              ||||||||||||   ||||||   ||||||||||||||||||||||||||||||||||||
Db         75 TTGTTGCCATATTTCCTGCCGATTCATCCAGGGCGAATGTGGTGTCATTGCCGGATGGTT 134

Qy        262 TyrLeuProMetSerTyrLeuTyrGlyLysArgPheValGlyProIleSerProThrVal 281
              ||||||||||||   |||:::||||||||||||||||||||||:::::::|||   :::
Db        135 TATTTGCCTATGTGTTACATTTATGGAAAGAGGTTTGTGGGCCCAGTTACACCAATTATA 194

Qy        282 LeuSerLeuArgLysGluLeuTyrThrValProTyrHisAspIleAspTrpAspGlnAla 301
              |||   ||||||||||||||||||   |||||||||:::::::||||||||:::|||
Db        195 TTGGAATTAAGAAAGGAACTCTACGAAGTACCCTACAATGAAGTTGATTGGGACAAGGCT 254

Qy        302 ArgAsnLeuCysAlaLysGluAspLeuTyrTyrProHisProLeuValGlnAspIleLeu 321
              ||||||||||||||||||||||||||||||||||||||||||||||   |||||||||:::|||
Db        255 CGCAATCTATGTGCTAAGGAAGATCTGTACTATCCACATCCATTCGTGCAGGATGTATTA 314

Qy        322 TrpAlaSerLeuHisLysPheLeuGluProIleLeuMetHisTrpProGlyLysArgLeu 341
              ||||||:::||||||||||||:::||||||   ::::::   |||||||||   :::|||
Db        315 TGGGCCACTCTCCACAAATTTGTTGAACCAGCTATGTTGCGTTGGCCTGGGAACAAATTG 374

Qy        342 ArgGluLysAlaIleIleSerAlaLeuGluHisIleHisTyrGluAspGluAsnThrArg 361
              ||||||||||||:::   :::   ::::::||||||||||||||||||||||||||||||
Db        375 AGGGAGAAAGCTTTGGACACTGTCATGCAGCATATTCATTATGAAGATGAGAACACCCGA 434

Qy        362 TyrIleCysIleGlyProValAsnLysValLeuAsnMetLeuCysCysTrpValGluAsp 381
              ||||||||||||||||||||||||||||||||||||||||||||   ||||||:::||||||
Db        435 TATATTTGCATTGGTCCAGTAAACAAGGTATTAAATATGCTTGCTTGCTGGATTGAAGAT 494

Qy        382 ProAsnSerGluAlaPheLysLeuHisLeuProArgIleTyrAspTyrLeuTrpIleAla 401
              ||||||||||||||||||||||||||:::||||||:::::||||||||||||||||||||
Db        495 CCAAACTCAGAGGCATTCAAACTCCACATTCCAAGAGTCCACGATTACCTATGGATTGCA 554

Qy        402 GluAspGlyMetLysMetGlnGlyTyrAsnGlySerGlnLeuTrpAspThrAlaPheAla 421
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        555 GAAGATGGCATGAAAATGCAGGGTTATAATGGAAGCCAGCTGTGGGACACAGCTTTCACA 614

Qy        422 ValGlnAlaIleIleAlaSerAsnLeuIleGluGluPheGlyProThrIleArgLysAla 441
              ||||||||||||:::|||:::   ||||||||||||||||||||||||||:::::   |||
Db        615 GTTCAAGCTATAGTGGCTACTGGCCTCATTGAAGAATTTGGTCCTACTCTTAAACTAGCA 674

Qy        442 HisThrTyrIleLysAsnSerGlnValLeuGluAspCysProGlyAspLeuAsnLysTrp 461
              |||   |||||||||   :::||||||:::::||||||||||||||||||:::::|||
Db        675 CATGGCTACATAAAGAAAACGCAGGTTATCGATGACTGCCCTGGAGATCTTAGTCAGTGG 734

Qy        462 TyrArgHisIleSerLysGlyAlaTrpProPheSerThrGlyAspHisGlyTrpProIle 481
              ||||||||||||||||||||||||||||||||||||||||||||   ||||||||||||||||
Db        735 TACCGCCACATATCTAAAGGTGCATGGCCCTTTTCTACTGCTGATCATGGTTGGCCTATA 794

Qy        482 SerAspCysThrAlaGluGlyLeuLysAlaValLeuLeuLeuSerLysIleAlaProGlu 501
              ||||||||||||||||||||||||||||||   ||||||||||||||||||:::|||:::
Db        795 TCAGATTGCACTGCAGAAGGACTTAAGGCGGCATTATTGCTATCGAAGATTTCTCCAGAT 854

Qy        502 IleValGlyGluProIleAspValLysArgLeuTyrAspSerValAsnValIleLeuSer 521
              ||||||||||||   :::::::|||   ||||||||||||||||||   ::::::|||
Db        855 ATTGTTGGCGAAGCAGTGGAAGTTAATAGACTGTATGATTCTGTCAATTGTTTGATGTCA 914

Qy        522 LeuGlnAsnGluAspGlyGlyPheAlaThrTyrGluLeuLysArgSerTyrAsnTrpLeu 541
                    |||::::::||||||||||||||||||||||||||||||   |||||||||   ||||||
Db        915 TACATGAATGATAATGGTGGATTTGCAACATATGAACTCACAAGGTCTTATGCCTGGCTG 974

Qy        542 GluIleIleAsnProAlaGluThrPheGlyAspIleValIleAspTyrProTyrValGlu 561
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        975 GAGCTTATCAATCCTGCTGAGACCTTTGGGGACATTGTGATTGATTATCCTTATGTGGAA 1034

Qy        562 CysThrSerAlaAlaIleGlnAlaLeuAlaSerPheArgLysLeuTyrProGlyHisArg 581
              ||||||||||||||||||||||||||||   :::|||:::||||||||||||||||||||
```

```
Db       2185 GAAATAATGGGAGTCTTCAACAAGAATTGCATGATCACATACGCCGCATACAGAAACATA 2244

Qy        741 PheProIleTrpAlaLeuGlyGluTyrGlnSerGlnValLeuGlnSerArg 757
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2245 TTCCCCATTTGGGCGTTGGGAGAATACCAATCCCAAGTATTGCAATCTCGT 2295
```

RESULT 2
US-09-419-679-1
; Sequence 1, Application US/09419679
; Patent No. 6630617
; GENERAL INFORMATION:
; APPLICANT: Falco, S. Carl
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Hitz, William D.
; APPLICANT: Kinney, Anthony J.
; APPLICANT: Rafalski, Antoni
; APPLICANT: McGonigle, Brian
; APPLICANT: Lohman, Karin
; TITLE OF INVENTION: Enzymes Involved In Squalene Metabolizm
; FILE REFERENCE: BB1113 US NA
; CURRENT APPLICATION NUMBER: US/09/419,679
; CURRENT FILING DATE: 1999-10-15
; EARLIER APPLICATION NUMBER: 60/105,405
; EARLIER FILING DATE: 1998-10-23
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 1
; LENGTH: 2558
; TYPE: DNA
; ORGANISM: Zea mays
US-09-419-679-1

Alignment Scores:

Pred. No.:	0	Length:	2558
Score:	3125.00	Matches:	545
Percent Similarity:	88.87%	Conservative:	78
Best Local Similarity:	77.75%	Mismatches:	78
Query Match:	75.34%	Indels:	0
DB:	4	Gaps:	0

US-09-419-679-6 (1-757) x US-09-419-679-1 (1-2558)

```
Qy         54 ThrHisLysHisSerAlaAspLeuLeuMetArgMetGlnPheAlaArgGluAsnProThr 73
              |||   ||||||||||||||||||||||||||:::|||||||||:::||||||
Db          3 ACGAGCAAGCACAGCGCCGACCTCCTCATGCGCATCCAGTTCGCCAAAGAAAACTCGATT 62

Qy         74 ArgGluValLeuProLysValGlyValLysAspIleGluAspValThrGlnGluIleVal 93
              ||||||   :::   :::   :::   ||||||||||||:::|||   |||
Db         63 GAGCTTCACCTTCCAGGCATCAAGCTCGGTGAGCATGAAGATGTGACCGAGGAAGCTGTG 122

Qy         94 ThrLysThrLeuArgArgAlaValSerPheHisSerThrLeuGlnCysHisAspGlyHis 113
              ||||||:::||||||:::|||       ||||||||||||  ||||||||||||
Db        123 TTGACTACATTGAAAAGGGCAATCAGCCGTTTCTCTACTCTCCAGGCACATGATGGACAC 182

Qy        114 TrpProGlyAspTyrGlyGlyProMetPheLeuMetProGlyLeuValIleThrLeuSer 133
              ||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db        183 TGGCCTGGGGATTATGGTGGTCCTATGTTCCTTATGCCAGGCTTGATCATAACATTGTAT 242

Qy        134 IleThrGlyAlaLeuAsnThrValLeuThrGluGluHisArgLysGluIleCysArgTyr 153
              :::|||||||||||||||||||||||:::   ||||||:::||||||||   ||||||
Db        243 GTGACTGGAGCACTAAACACTGTCTTGTCATTGGAACACCAGAAGGAGATCCGCCGGTAT 302

Qy        154 LeuTyrAsnHisGlnAsnLysAspGlyGlyTrpGlyLeuHisIleGluGlyProSerThr 173
              ||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db        303 CTTTATAATCACCAGAATGAAGATGGCGGCTGGGGCTTGCACATTGAGGGTCCAAGCACC 362

Qy        174 MetPheGlySerValLeuSerTyrIleThrLeuArgLeuLeuGlyGluGlyProAsnAsp 193
              ||||||||||||   |||:::|||:::   ||||||||||||||||||||||:::
Db        363 ATGTTCGGCTCAGCACTGACCTATGTTATTTTGAGATTGCTTGGAGAGGGACCAGATAGT 422

Qy        194 GlyGlnGlyGluMetGluLysAlaArgAspTrpIleLeuGlyHisGlyGlyAlaThrTyr 213
              |||   |||   |||||||||   |||:::|||||||||   ||||||||||||||||||
Db        423 GGAGATGGAGCCATGGAGAAAGGTCGAAACTGGATATTAGACCATGGTGGAGCAACATAT 482

Qy        214 IleThrSerTrpGlyLysMetTrpLeuSerValLeuGlyValTyrGluTrpSerGlyAsn 233
              ||||||||||||||||||   ||||||||||||||||||||:::|||||||||||||||
Db        483 ATAACATCGTGGGGGAAGTTTTGGCTTTCGGTACTAGGTGTATTTGAATGGTCTGGTAAC 542

Qy        234 AsnProLeuProProGluIleTrpLeuLeuProTyrMetLeuProPheHisProGlyArg 253
              ||||||:::||||||||||:::||||||||||||||||:::||||||||||||||||||||
Db        543 AACCCGGTGCCACCAGAAGTATGGCTACTGCCATATCTCCTCCCATTTCACCCAGGGAGG 602

Qy        254 MetTrpCysHisCysArgMetValTyrLeuProMetSerTyrLeuTyrGlyLysArgPhe 273
              ||||||||||||||||||||||||||||||||||||||||   |||:::|||||||||||||
Db        603 ATGTGGTGTCATTGTCGAATGGTGTATTTGCCAATGTGCTACATTTATGGGAAGAGGTTT 662

Qy        274 ValGlyProIleSerProThrValLeuSerLeuArgLysGluLeuTyrThrValProTyr 293
              ||||||   |||:::|||   :::||||   ||||||||||||||:::      ||||||
Db        663 GTTGGCCGAATCACACCACTTCTGTTGGAATTAAGGAAGGAACTTTTCAAAGACCCTTAC 722

Qy        294 HisAspIleAspTrpAspGlnAlaArgAsnLeuCysAlaLysGluAspLeuTyrTyrPro 313
                 ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||
Db        723 AGCAAGATTGATTGGGACAAGGCCCGCAACCTATGTGCCAAGGAAGATCTGTACTACCCA 782

Qy        314 HisProLeuValGlnAspIleLeuTrpAlaSerLeuHisLysPheLeuGluProIleLeu 333
              ||||||   |||||||||:::||||||||:::||||||||||||:::||||||::::::
Db        783 CACCCATTCGTTCAAGATGTGTTGTGGGCCACTCTCCATAAATTCGTTGAACCAGTTATG 842

Qy        334 MetHisTrpProGlyLysArgLeuArgGluLysAlaIleIleSerAlaLeuGluHisIle 353
              ||||||||||||||   :::||||||||||||||:::   :::   ::::::|||:::
Db        843 ATGCATTGGCCTGGCAGCAAATTGAGGGAGAAAGCTCTGGAAACAGTCATGCAACATGTT 902

Qy        354 HisTyrGluAspGluAsnThrArgTyrIleCysIleGlyProValAsnLysValLeuAsn 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        903 CATTATGAAGATGAGAACACTCGTTATATTTGCATTGGTCCTGTAAACAAGGTATTGAAT 962

Qy        374 MetLeuCysCysTrpValGluAspProAsnSerGluAlaPheLysLeuHisLeuProArg 393
              ||||||   ||||||:::||||||||||||||||||||||||||||||:::||||||
Db        963 ATGCTTGCTTGCTGGATTGAAGATCCAAACTCGGAGGCCTTCAAACTTCATATCCCACGA 1022

Qy        394 IleTyrAspTyrLeuTrpIleAlaGluAspGlyMetLysMetGlnGlyTyrAsnGlySer 413
              :::||||||||||||||:::||||||||||||||||||||||||||||||||||||||||
Db       1023 GTCTATGATTACTTGTGGCTTGCTGAAGATGGCATGAAGATGCAGGGTTATAATGGCAGC 1082

Qy        414 GlnLeuTrpAspThrAlaPheAlaValGlnAlaIleIleAlaSerAsnLeuIleGluGlu 433
              ||||||||||||||||||||   ||||||||||||:::|||:::||||||||||||||||
Db       1083 CAACTTTGGGATACAGCTTTCACAGTTCAAGCCATTGTGGCTACCAACCTTATTGAAGAG 1142

Qy        434 PheGlyProThrIleArgLysAlaHisThrTyrIleLysAsnSerGlnValLeuGluAsp 453
              ||||||||||||:::::   ||||||   ||||||||||||||||||||||:::|||
Db       1143 TTTGGTCCTACCCTTAAACTAGCACACAACTATATCAAGAATTCACAGGTTCTTGATGAC 1202

Qy        454 CysProGlyAspLeuAsnLysTrpTyrArgHisIleSerLysGlyAlaTrpProPheSer 473
              ||||||||||||||||||   ||||||||||   ||||||||||||||||||||||||||
Db       1203 TGCCCTGGTGATCTGAATGACTGGTACCGCCACACATCTAAAGGTGCATGGCCATTCTCA 1262

Qy        474 ThrGlyAspHisGlyTrpProIleSerAspCysThrAlaGluGlyLeuLysAlaValLeu 493
              |||   ||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db       1263 ACTGCTGATCATGGTTGGCCTATATCTGATTGCACTGCTGAAGGACTAAAGGCTTCATTA 1322

Qy        494 LeuLeuSerLysIleAlaProGluIleValGlyGluProIleAspValLysArgLeuTyr 513
              |||||||||:::|||:::|||:::||||||||||||||:::::   |||   |||
Db       1323 TTGTTATCAAGGATCTCTCCCAAAATTGTTGGTGAACCGATGGAAGCTAATAGATTTTAT 1382

Qy        514 AspSerValAsnValIleLeuSerLeuGlnAsnGluAspGlyGlyPheAlaThrTyrGlu 533
              |||:::|||:::   ::::::|||       |||::::::||||||||||||||||||||
Db       1383 GATGCTGTCAGTTGTCTGATGTCTTATATGAATGATAATGGCGGTTTCGCGACATATGAA 1442

Qy        534 LeuLysArgSerTyrAsnTrpLeuGluIleIleAsnProAlaGluThrPheGlyAspIle 553
              |||   |||||||||   |||||||||:::||||||||||||||||||||||||||||||
Db       1443 CTCACAAGATCTTATCCCTGGTTGGAGCTGATCAATCCCGCTGAGACCTTTGGGGATATT 1502

Qy        554 ValIleAspTyrProTyrValGluCysThrSerAlaAlaIleGlnAlaLeuAlaSerPhe 573
              ||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db       1503 GTGATTGATTACCCGTATGTTGAATGTACATCAGCAGCAATTCAGGCCCTGACATCATTC 1562
```

```
Qy          1 MetTrpLysLeuLysPheAlaGluGlyGlyAsnProTrpLeuArgThrLeuAsnAsnHis 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         25 ATGTGGAAGCTCAAGTTCGCCGAAGGAGGGAATCCATGGCTTCGGACATTGAACAATCAC 84

Qy         21 ValGlyArgGlnValTrpGluPheAspProLysLeuGlySerProGlnAspLeuLeuGlu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         85 GTTGGAAGACAGGTGTGGGAGTTCGATCCTAAGCTTGGATCGCCGCAAGATCTCCTCGAG 144

Qy         41 IleGluLysAlaArgGlnAsnPheHisAspAsnArgPheThrHisLysHisSerAlaAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        145 ATTGAGAAAGCTCGCCAGAATTTTCACGATAACCGCTTTACCCACAAACACAGCGCTGAT 204

Qy         61 LeuLeuMetArgMetGlnPheAlaArgGluAsnProThrArgGluValLeuProLysVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        205 CTACTTATGCGGATGCAGTTCGCAAGAGAGAACCCAACACGTGAAGTCTTGCCCAAAGTC 264

Qy         81 GlyValLysAspIleGluAspValThrGlnGluIleValThrLysThrLeuArgArgAla 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        265 GGAGTTAAGGATATTGAGGATGTGACCCAAGAGATTGTGACAAAAACATTAAGAAGGGCC 324

Qy        101 ValSerPheHisSerThrLeuGlnCysHisAspGlyHisTrpProGlyAspTyrGlyGly 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        325 GTAAGTTTCCATTCAACTCTCCAGTGCCATGACGGACACTGGCCGGGAGATTATGGAGGT 384

Qy        121 ProMetPheLeuMetProGlyLeuValIleThrLeuSerIleThrGlyAlaLeuAsnThr 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        385 CCCATGTTTCTGATGCCTGGCTTGGTAATTACTCTGTCTATCACTGGGGCGTTGAATACA 444

Qy        141 ValLeuThrGluGluHisArgLysGluIleCysArgTyrLeuTyrAsnHisGlnAsnLys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        445 GTCTTAACTGAAGAACATAGAAAGGAAATATGCCGTTACCTCTATAATCATCAAAACAAG 504

Qy        161 AspGlyGlyTrpGlyLeuHisIleGluGlyProSerThrMetPheGlySerValLeuSer 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        505 GATGGTGGGTGGGGTTTGCATATTGAAGGTCCAAGCACCATGTTTGGCTCTGTCTTGAGT 564

Qy        181 TyrIleThrLeuArgLeuLeuGlyGluGlyProAsnAspGlyGlnGlyGluMetGluLys 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        565 TATATTACTCTGAGATTGCTAGGTGAGGGGCCTAATGATGGACAAGGGGAAATGGAGAAG 624

Qy        201 AlaArgAspTrpIleLeuGlyHisGlyGlyAlaThrTyrIleThrSerTrpGlyLysMet 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        625 GCACGTGACTGGATTCTAGGGCATGGTGGTGCTACTTATATAACGTCATGGGGGAAGATG 684

Qy        221 TrpLeuSerValLeuGlyValTyrGluTrpSerGlyAsnAsnProLeuProProGluIle 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        685 TGGCTTTCAGTACTTGGAGTGTATGAATGGTCTGGAAATAATCCCCTGCCCCCTGAGATA 744

Qy        241 TrpLeuLeuProTyrMetLeuProPheHisProGlyArgMetTrpCysHisCysArgMet 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        745 TGGCTCCTTCCATACATGCTTCCATTTCATCCAGGAAGGATGTGGTGTCACTGCCGGATG 804

Qy        261 ValTyrLeuProMetSerTyrLeuTyrGlyLysArgPheValGlyProIleSerProThr 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        805 GTCTATTTGCCGATGTCCTACTTATATGGCAAGAGGTTTGTTGGTCCAATCTCACCAACA 864

Qy        281 ValLeuSerLeuArgLysGluLeuTyrThrValProTyrHisAspIleAspTrpAspGln 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        865 GTATTATCTTTGAGAAAAGAGCTTTATACAGTACCATACCATGATATAGATTGGGATCAG 924

Qy        301 AlaArgAsnLeuCysAlaLysGluAspLeuTyrTyrProHisProLeuValGlnAspIle 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        925 GCTCGCAATTTGTGTGCAAAGGAAGATTTGTACTATCCTCACCCACTTGTACAGGATATT 984

Qy        321 LeuTrpAlaSerLeuHisLysPheLeuGluProIleLeuMetHisTrpProGlyLysArg 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        985 CTTTGGGCATCTCTACACAAGTTCCTTGAGCCTATTCTGATGCATTGGCCTGGAAAAAGA 1044

Qy        341 LeuArgGluLysAlaIleIleSerAlaLeuGluHisIleHisTyrGluAspGluAsnThr 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1045 TTGAGGGAAAAGGCTATTATTTCTGCATTGGAGCATATACATTACGAAGATGAGAATACT 1104

Qy        361 ArgTyrIleCysIleGlyProValAsnLysValLeuAsnMetLeuCysCysTrpValGlu 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1105 CGATATATTTGCATAGGTCCTGTAAATAAGGTGTTAAATATGCTTTGCTGTTGGGTGGAA 1164

Qy        381 AspProAsnSerGluAlaPheLysLeuHisLeuProArgIleTyrAspTyrLeuTrpIle 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1165 GATCCAAATTCTGAGGCCTTCAAGTTGCATCTTCCCAGGATTTATGATTATCTATGGATT 1224

Qy        401 AlaGluAspGlyMetLysMetGlnGlyTyrAsnGlySerGlnLeuTrpAspThrAlaPhe 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1225 GCAGAAGATGGCATGAAAATGCAGGGCTACAATGGAAGTCAACTATGGGACACTGCTTTT 1284

Qy        421 AlaValGlnAlaIleIleAlaSerAsnLeuIleGluGluPheGlyProThrIleArgLys 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1285 GCTGTCCAAGCAATTATTGCATCTAACCTCATTGAAGAATTTGGTCCAACTATAAGAAAA 1344

Qy        441 AlaHisThrTyrIleLysAsnSerGlnValLeuGluAspCysProGlyAspLeuAsnLys 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1345 GCTCATACCTATATTAAGAATTCACAGGTTTTAGAAGATTGTCCAGGTGATCTTAATAAA 1404

Qy        461 TrpTyrArgHisIleSerLysGlyAlaTrpProPheSerThrGlyAspHisGlyTrpPro 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1405 TGGTACCGTCACATTTCAAAAGGTGCTTGGCCTTTTTCAACTGGAGATCATGGATGGCCA 1464

Qy        481 IleSerAspCysThrAlaGluGlyLeuLysAlaValLeuLeuLeuSerLysIleAlaPro 500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1465 ATTTCTGACTGCACAGCTGAAGGACTGAAAGCTGTTCTATTACTATCCAAAATTGCACCA 1524

Qy        501 GluIleValGlyGluProIleAspValLysArgLeuTyrAspSerValAsnValIleLeu 520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1525 GAAATAGTTGGTGAGCCAATAGACGTGAAGCGATTATATGATTCTGTAAATGTCATTCTC 1584

Qy        521 SerLeuGlnAsnGluAspGlyGlyPheAlaThrTyrGluLeuLysArgSerTyrAsnTrp 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1585 TCACTACAGAATGAAGATGGTGGTTTTGCAACATATGAGCTTAAACGATCTTATAATTGG 1644

Qy        541 LeuGluIleIleAsnProAlaGluThrPheGlyAspIleValIleAspTyrProTyrVal 560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1645 TTGGAGATAATCAATCCTGCTGAAACTTTTGGTGACATCGTTATTGATTATCCTTATGTG 1704

Qy        561 GluCysThrSerAlaAlaIleGlnAlaLeuAlaSerPheArgLysLeuTyrProGlyHis 580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1705 GAATGTACATCAGCAGCGATTCAAGCTTTGGCATCATTTAGGAAATTATATCCTGGGCAT 1764

Qy        581 ArgArgGluGluIleGlnHisCysIleAspLysAlaThrThrPheIleGluLysIleGln 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1765 CGCCGAGAAGAAATACAACATTGTATCGATAAAGCCACTACCTTCATTGAAAAAATACAA 1824

Qy        601 AlaSerAspGlySerTrpTyrGlySerTrpGlyValCysPheThrTyrGlyAlaTrpPhe 620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1825 GCTTCAGATGGATCATGGTATGGTTCTTGGGGAGTTTGCTTCACTTACGGTGCTTGGTTT 1884

Qy        621 GlyValLysGlyLeuIleAlaAlaGlyArgSerPheSerAsnCysSerSerIleArgLys 640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1885 GGGGTAAAAGGTCTGATTGCTGCTGGAAGGAGTTTCAGTAATTGCTCAAGCATCCGTAAA 1944

Qy        641 AlaCysGluPheLeuLeuSerLysGlnLeuProSerGlyGlyTrpGlyGluSerTyrLeu 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1945 GCTTGTGAATTTCTGCTGTCCAAGCAGCTTCCTTCTGGTGGCTGGGGAGAGAGTTATCTG 2004

Qy        661 SerCysGlnAsnLysValTyrSerAsnLeuGluGlyAsnArgSerHisValValAsnThr 680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2005 TCCTGTCAAAACAAGGTGTATTCAAATCTGGAAGGCAACAGGTCTCATGTGGTCAACACT 2064

Qy        681 GlyTrpAlaMetLeuAlaLeuIleAspAlaGlyGlnAlaLysArgAspSerGlnProLeu 700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2065 GGGTGGGCTATGTTGGCTCTCATTGATGCTGGACAGGCTAAGAGAGATTCGCAACCACTG 2124

Qy        701 HisArgAlaAlaAlaTyrLeuIleAsnSerGlnLeuGluAspGlyAspProProGlnGln 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2125 CACCGGGCAGCTGCATACTTGATAAATTCCCAATTGGAGGACGGTGACTTTCCGCAGCAG 2184

Qy        721 GluIleMetGlyValPheAsnLysAsnCysMetIleThrTyrAlaAlaTyrArgAsnIle 740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 5, 2005, 12:17:05 ; Search time 328 Seconds
(without alignments)
3776.407 Million cell updates/sec

Title: US-09-419-679-6
Perfect score: 4148
Sequence: 1 MWKLKFAEGGNPWLRTLNNH.....RNIFPIWALGEYQSQVLQSR 757

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/6630617/runat_02112005_120302_15105/app_query.fasta_1.903
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=6630617 @CGN_1_1_69_@runat_02112005_120302_15105 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

from patent application

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4148	100.0	2560	4	US-09-419-679-5	Sequence 5, Appli
2	3125	75.3	2558	4	US-09-419-679-1	Sequence 1, Appli
3	2433	58.7	1882	4	US-09-419-679-3	Sequence 3, Appli
4	1653.5	39.9	4346	4	US-09-919-039-112	Sequence 112, App
5	1597	38.5	4239	4	US-09-815-048-1	Sequence 1, Appli
6	1442	34.8	1300	4	US-09-419-679-7	Sequence 7, Appli
7	1276	30.8	2196	4	US-09-614-221A-308	Sequence 308, App
8	403.5	9.7	747	4	US-09-248-796A-3811	Sequence 3811, Ap
c 9	361	8.7	536165	4	US-09-214-808-1	Sequence 1, Appli
c 10	340.5	8.2	539	3	US-08-998-416-229	Sequence 229, App
11	198	4.8	2072	4	US-09-725-735A-11	Sequence 11, Appl
12	165	4.0	173	4	US-09-313-294A-804	Sequence 804, App
13	158	3.8	42450	4	US-09-815-048-3	Sequence 3, Appli
14	151	3.6	255	4	US-09-248-796A-3810	Sequence 3810, Ap
15	145.5	3.5	5046	4	US-09-725-735A-13	Sequence 13, Appl
c 16	121	2.9	33140	4	US-09-596-002-23	Sequence 23, Appl
17	118.5	2.9	1401	4	US-09-583-110-684	Sequence 684, App
18	118.5	2.9	1419	4	US-09-107-433-1867	Sequence 1867, Ap
19	116	2.8	1306	3	US-09-387-574-3	Sequence 3, Appli
20	116	2.8	1306	3	US-09-668-096-3	Sequence 3, Appli
c 21	114.5	2.8	16047	4	US-09-902-540-1136	Sequence 1136, Ap
c 22	113.5	2.7	7939	3	US-08-961-527-9	Sequence 9, Appli
23	113.5	2.7	1830121	4	US-09-557-884-1	Sequence 1, Appli
24	113.5	2.7	1830121	4	US-09-643-990A-1	Sequence 1, Appli
25	113	2.7	23451	3	US-09-453-702B-173	Sequence 173, App
c 26	112.5	2.7	129908	4	US-09-585-858-1	Sequence 1, Appli
c 27	112.5	2.7	129908	4	US-10-270-878-1	Sequence 1, Appli
28	111	2.7	23901	4	US-09-949-016-16773	Sequence 16773, A
29	110.5	2.7	10474	4	US-09-596-141C-9	Sequence 9, Appli
30	110.5	2.7	10474	4	US-09-595-526C-9	Sequence 9, Appli
31	110	2.7	2658	3	US-08-727-308-2	Sequence 2, Appli
32	110	2.7	2658	3	US-08-727-308-3	Sequence 3, Appli
c 33	110	2.7	26502	4	US-09-949-016-14585	Sequence 14585, A
34	109.5	2.6	7860	4	US-09-526-193A-2	Sequence 2, Appli
35	109.5	2.6	12284	2	US-08-876-991-1	Sequence 1, Appli
36	109.5	2.6	12284	2	US-09-059-853-1	Sequence 1, Appli
37	109	2.6	111282	3	US-09-754-250-3	Sequence 3, Appli
38	108	2.6	14335	4	US-09-596-002-11	Sequence 11, Appl
39	107.5	2.6	7497	4	US-09-949-016-3114	Sequence 3114, Ap
c 40	107.5	2.6	57280	4	US-09-949-016-11796	Sequence 11796, A
c 41	107.5	2.6	57280	4	US-09-949-016-12843	Sequence 12843, A
c 42	107.5	2.6	57280	4	US-09-949-016-12844	Sequence 12844, A
c 43	107.5	2.6	57280	4	US-09-949-016-12846	Sequence 12846, A
c 44	107.5	2.6	57280	4	US-09-949-016-13542	Sequence 13542, A
c 45	107.5	2.6	57280	4	US-09-949-016-13543	Sequence 13543, A

ALIGNMENTS

RESULT 1
US-09-419-679-5
; Sequence 5, Application US/09419679
; Patent No. 6630617
; GENERAL INFORMATION:
; APPLICANT: Falco, S. Carl
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Hitz, William D.
; APPLICANT: Kinney, Anthony J.
; APPLICANT: Rafalski, Antoni
; APPLICANT: McGonigle, Brian
; APPLICANT: Lohman, Karin
; TITLE OF INVENTION: Enzymes Involved In Squalene Metabolizm
; FILE REFERENCE: BB1113 US NA
; CURRENT APPLICATION NUMBER: US/09/419,679
; CURRENT FILING DATE: 1999-10-15
; EARLIER APPLICATION NUMBER: 60/105,405
; EARLIER FILING DATE: 1998-10-23
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 5
; LENGTH: 2560
; TYPE: DNA
; ORGANISM: Glycine max
US-09-419-679-5

Alignment Scores:

Pred. No.:	0	Length:	2560
Score:	4148.00	Matches:	757
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	4	Gaps:	0

US-09-419-679-6 (1-757) x US-09-419-679-5 (1-2560)

```
  Query Match             1.7%;  Score 43.6;  DB 4;  Length 403;
  Best Local Similarity   49.3%;  Pred. No. 0.0058;
  Matches  140;  Conservative    1;  Mismatches  140;  Indels    3;  Gaps    1;

Qy       1409 ACCGTCACATTTCAAAAGGTGCTTGGCCTTTTTCAACTGGAGATCATGGATGGCCAATTT 1468
              ||: ||||  | | |||| ||   |  ||    ||||   ||||||         ||
Db         44 ACMATCACTCTACTAAAGTTGTMAGTACTGACACAACAAAAGATCAAACTAAAACACAAA 103

Qy       1469 CTGACTGCACAGCTGAAGGACTGAAAGCTGTTCTATTACTATCCAAAATT---GCACCAG 1525
              |||     |||| | ||  |    || ||| || |  || |   ||| ||    |||| |
Db        104 CTGCTCATACAGTTAAAACAGCACAAACTGCTCAAGAACAAAATAAAGTTCAAACACCTG 163

Qy       1526 AAATAGTTGGTGAGCCAATAGACGTGAAGCGATTATATGATTCTGTAAATGTCATTCTCT 1585
                | || || ||   ||     |      |  |      || |  |||||| ||  | |   |
Db        164 TTAAAGATGTTGCAACAGCGAAATCTGAAAGCAACAATCAAGCTGTAAGTGATAATAAAT 223

Qy       1586 CACTACAGAATGAAGATGGTGGTTTTGCAACATATGAGCTTAAACGATCTTATAATTGGT 1645
              ||| ||| | | |  | | |          |   |   || ||||| | | | |||
Db        224 CACAACAAACTAACAAAGTTACAAAACATAACGAAACGCCTAAACAAGCATCTAAAGCTA 283

Qy       1646 TGGAGATAATCAATCCTGCTGAAACTTTTGGTGACATCGTTATT 1689
                ||  ||   ||  ||| |  |||||  | ||| | | |||||
Db        284 AAGAATTACCAAAAACTGGTTTAACTTCAGTTGATAACTTTATT 327


RESULT 15
US-08-781-986A-767
; Sequence 767, Application US/08781986A
; Patent No. 6737248
;  GENERAL INFORMATION:
;    APPLICANT:  Charles Kunsch
;    TITLE OF INVENTION:  Staphylococcus aureus Polynucleotides and Sequences
;    NUMBER OF SEQUENCES:  5255
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/781,986A
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Benson, Bob
;      REGISTRATION NUMBER:  30,446
;      REFERENCE/DOCKET NUMBER:  PB248PP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  767:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  403 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-781-986A-767

  Query Match             1.7%;  Score 43.6;  DB 4;  Length 403;
  Best Local Similarity   49.3%;  Pred. No. 0.0058;
  Matches  140;  Conservative    1;  Mismatches  140;  Indels    3;  Gaps    1;

Qy       1409 ACCGTCACATTTCAAAAGGTGCTTGGCCTTTTTCAACTGGAGATCATGGATGGCCAATTT 1468
              ||: ||||  | | |||| ||   |  ||    ||||   ||||||         ||
Db         44 ACMATCACTCTACTAAAGTTGTMAGTACTGACACAACAAAAGATCAAACTAAAACACAAA 103

Qy       1469 CTGACTGCACAGCTGAAGGACTGAAAGCTGTTCTATTACTATCCAAAATT---GCACCAG 1525
              |||     |||| | ||  |    || ||| || |  || |   ||| ||    |||| |
Db        104 CTGCTCATACAGTTAAAACAGCACAAACTGCTCAAGAACAAAATAAAGTTCAAACACCTG 163

Qy       1526 AAATAGTTGGTGAGCCAATAGACGTGAAGCGATTATATGATTCTGTAAATGTCATTCTCT 1585
                | || || ||   ||     |      |  |      || |  |||||| ||  | |   |
Db        164 TTAAAGATGTTGCAACAGCGAAATCTGAAAGCAACAATCAAGCTGTAAGTGATAATAAAT 223

Qy       1586 CACTACAGAATGAAGATGGTGGTTTTGCAACATATGAGCTTAAACGATCTTATAATTGGT 1645
              ||| ||| | | |  | | |          |   |   || ||||| | | | |||
Db        224 CACAACAAACTAACAAAGTTACAAAACATAACGAAACGCCTAAACAAGCATCTAAAGCTA 283

Qy       1646 TGGAGATAATCAATCCTGCTGAAACTTTTGGTGACATCGTTATT 1689
                ||  ||   ||  ||| |  |||||  | ||| | | |||||
Db        284 AAGAATTACCAAAAACTGGTTTAACTTCAGTTGATAACTTTATT 327


Search completed: November  5, 2005, 12:16:35
Job time : 440 secs
```

; NAME/KEY: misc_feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g

Query Match 1.7%; Score 44; DB 4; Length 1664976;

Best Local Similarity 46.9%; Pred. No. 1.7;
Matches 137; Conservative 0; Mismatches 155; Indels 0; Gaps 0;

```
Qy       1118 TAGGTCCTGTAAATAAGGTGTTAAATATGCTTTGCTGTTGGGTGGAAGATCCAAATTCTG 1177
              ||  | || ||||          ||   | | | |      ||  || |||      ||  |
Db    1268220 TAAATTCTATAAAACTTAGTGCAATAGTTCATGCCACAGAGGATGAGGATAAGGTTTTAG 1268161

Qy       1178 AGGCCTTCAAGTTGCATCTTCCCAGGATTTATGATTATCTATGGATTGCAGAAGATGGCA 1237
              ||||  |  | ||   | ||||   ||     |||| |   |   || |    |||||
Db    1268160 AGGCGATAGAATTTTTTATTCCTGAGAACGTTGATGAGGAAAAAATAGACTTAGATGTTG 1268101

Qy       1238 TGAAAATGCAGGGCTACAATGGAAGTCAACTATGGGACACTGCTTTTGCTGTCCAAGCAA 1297
              |  ||| |||||| |||  |||||  | | ||       |    | |   |||  ||| ||
Db    1268100 TTGAAACGCAGGGATACTTTGGAAACCCAATAAAAATTATCAATGTCAATGTTGAAGGAA 1268041

Qy       1298 TTATTGCATCTAACCTCATTGAAGAATTTGGTCCAACTATAAGAAAAGCTCATACCTATA 1357
                ||    ||  |  || || |  ||          |||| |  || | |||   |||
Db    1268040 AAGAGGCTAAGAAGATATTTAAACATATTATAGATTTAATAAAATCAGATGATAAAAATA 1267981

Qy       1358 TTAAGAATTCACAGGTTTTAGAAGATTGTCCAGGTGATCTTAATAAATGGTA 1409
              | || ||    ||     | | |   || |||  ||| |||| ||
Db    1267980 TAAATAAACTGAAGAAAGATTTACACTTAAGAGTTGAGGATAACAAATTTTA 1267929
```

RESULT 14
US-08-956-171E-767
; Sequence 767, Application US/08956171E
; Patent No. 6593114
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; Gil H. Choi
; Patrick S. Dillon
; Craig A. Rosen
; Steven C. Barash
; Michael R. Fannon
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5256
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 indh, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956,171E
; FILING DATE: 20-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/009,861
; FILING DATE: January 5, 1996
; APPLICATION NUMBER: 08/781,986
; FILING DATE: January 3, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mark J. Hyman
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB248P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (240) 314-1224
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 767:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 403 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 767:
US-08-956-171E-767

```
Db    1268160 AGGCGATAGAATTTTTTATTCCTGAGAACGTTGATGAGGAAAAAATAGACTTAGATGTTG 1268101

Qy       1238 TGAAAATGCAGGGCTACAATGGAAGTCAACTATGGGACACTGCTTTTGCTGTCCAAGCAA 1297

Db    1268100 TTGAAACGCAGGGATACTTTGGAAACCCAATAAAAATTATCAATGTCAATGTTGAAGGAA 1268041

Qy       1298 TTATTGCATCTAACCTCATTGAAGAATTTGGTCCAACTATAAGAAAAGCTCATACCTATA 1357

Db    1268040 AAGAGGCTAAGAAGATATTTAAACATATTATAGATTTAATAAAATCAGATGATAAAAATA 1267981

Qy       1358 TTAAGAATTCACAGGTTTTAGAAGATTGTCCAGGTGATCTTAATAAATGGTA 1409

Db    1267980 TAAATAAACTGAAGAAAGATTTACACTTAAGAGTTGAGGATAACAAATTTTA 1267929


RESULT 13
US-09-692-570-1/c
; Sequence 1, Application US/09692570
; Patent No. 6797466
; GENERAL INFORMATION:
;  APPLICANT: Bult et al.
;  TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococ
; Patent No. 6797466
;  TITLE OF INVENTION:  jannaschii
;  FILE REFERENCE: PB275C1
;  CURRENT APPLICATION NUMBER: US/09/692,570
;  CURRENT FILING DATE:  2003-01-14
;  PRIOR APPLICATION NUMBER: US 60/024,428
;  PRIOR FILING DATE: 1996-08-22
;  PRIOR APPLICATION NUMBER: US 08/916,421
;  PRIOR FILING DATE: 1997-08-22
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 1664976
;   TYPE: DNA
;   ORGANISM: Methanococcus jannaschii
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (28222)..(28222)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (28257)..(28258)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (84773)..(84773)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (84808)..(84808)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (84812)..(84812)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98120)..(98120)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98159)..(98159)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98239)..(98239)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98266)..(98266)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98343)..(98343)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (103998)..(103998)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (148948)..(148948)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (163385)..(163385)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (191989)..(191989)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (191995)..(191995)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (231980)..(231980)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (234187)..(234187)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (234220)..(234220)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (234814)..(234814)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (309398)..(309398)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (309418)..(309418)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (312837)..(312837)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (312993)..(312993)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (319226)..(319226)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (559167)..(559167)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (559241)..(559241)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (600992)..(600992)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
```

; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1603734)..(1603734)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1664854)..(1664855)
; OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

Query Match 1.7%; Score 44; DB 4; Length 1664976;
Best Local Similarity 46.9%; Pred. No. 1.7;
Matches 137; Conservative 0; Mismatches 155; Indels 0; Gaps 0;

```
Qy       1118 TAGGTCCTGTAAATAAGGTGTTAAATATGCTTTGCTGTTGGGTGGAAGATCCAAATTCTG 1177
              ||  | || ||||          ||   | | |  |      ||  || |||     || | |
Db    1268220 TAAATTCTATAAAACTTAGTGCAATAGTTCATGCCACAGAGGATGAGGATAAGGTTTTAG 1268161

Qy       1178 AGGCCTTCAAGTTGCATCTTCCCAGGATTTATGATTATCTATGGATTGCAGAAGATGGCA 1237
              ||||  |  | ||   | ||||  ||    |||| |   |   || |    |||||
```

```
  Query Match             2.3%;  Score 58.2;  DB 4;  Length 173;
  Best Local Similarity   63.0%;  Pred. No. 7.6e-08;
  Matches  109;  Conservative    0;  Mismatches   58;  Indels    6;  Gaps    1;

Qy       1485 AGGACTGAAAGCTGTTCTATTACTATCCAAAATTGCACCAGAAA------TAGTTGGTGA 1538
              || ||| || ||||| || || || || || || ||   ||     |||       | ||||| ||
Db          1 AGCACTTAAGGCTGTACTGTTGCTGTCAAAGATCTCAAACAAAAACAACCTTGTTGGGGA 60

Qy       1539 GCCAATAGACGTGAAGCGATTATATGATTCTGTAAATGTCATTCTCTCACTACAGAATGA 1598
              |||||||||      | ||  ||  ||||  || |  ||  | |||| ||  |   |||  |
Db         61 TCCAATAGAAAGAGAAAGGTTGCATGACGCTATTGATTGCCTTCTATCTTTTGCGAACAA 120

Qy       1599 AGATGGTGGTTTTGCAACATATGAGCTTAAACGATCTTATAATTGGTTGGAGA 1651
              ||||||    ||| | |||||||||   ||| || |||||  |||||| ||||
Db        121 AGATGGCACCTTTTCTACATATGAGTGCAAAAGAACTTATTCTTGGTTAGAGA 173


RESULT 11
US-09-806-708B-22/c
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
;  APPLICANT: The University of British Columbia
;  TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
;  FILE REFERENCE: 4810-58741
;  CURRENT APPLICATION NUMBER: US/09/806,708B
;  CURRENT FILING DATE:  2001-04-03
;  PRIOR APPLICATION NUMBER: US 60/147,133
;  PRIOR FILING DATE: 1999-08-04
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
;   LENGTH: 1141
;   TYPE: DNA
;   ORGANISM: Artificial sequence
;   FEATURE:
;   NAME/KEY: promoter
;   LOCATION: (1)..(1141)
;   OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22

  Query Match             1.8%;  Score 46;  DB 4;  Length 1141;
  Best Local Similarity   11.5%;  Pred. No. 0.0021;
  Matches   95;  Conservative  310;  Mismatches  410;  Indels   10;  Gaps    4;

Qy       1743 TAGGAAATTATATCCTGGGCATCGCCGAGAAGAAATACAACATTGTATCGATAAAGCCAC 1802
              :||  :: :  |   : | :| :  :|  :|: | ::|       | :::: |:
Db        980 WAGNNMRMYGAAAGNKWGCMAAMATMGBWWADTAGKMCNNNNNNWTTDVRRMAMKAKNNN 921

Qy       1803 TACCTTCATTGAAAAAATACAAGCTTCAGATGGATCATGGTATGGTTCTTGGGGAGTTTG 1862
                 :: | : :||   ::  : ::: :||::::  ::     ::|::       |:| :
Db        920 NNNAYWTACYNRAATNNKMATHWMKWTHGAHSKRRTRHHTRTCRRTKYNNNNNNARTVYW 861

Qy       1863 CTTCACTTACGGTGCTTGGTTTGGGGTAAAAGGTCTGATTG--CTGCTGGAAGGAGTTTC 1920
              ::: |  ::   :  |           :  : :  :::::                  |:|
Db        860 YHHAARRWMNAWWTRTNNNNNNNNNNACRNTRTWWABWKHSWCNNNNNNNNNNNNNNTWC 801

Qy       1921 AGTAATTGCTCAAGCATCCGTAAAGCTTGTGAATTTCTGCTGTCCAAGCAGCTTCCTTCT 1980
              : | |  :: ::|       :|: |     || || :  : :  ::  :  : : | :|
Db        800 HYTTANABBCYRANNNNAAARMARTCNNYMHAAVTTTHTDWCYKTWMNTWYWDMMTTMBT 741

Qy       1981 GGTGGCTGGGGAGAGAGTTAT----CTGTCCTGTCAAAACAAGGTGTATTCAAATCTGGA 2036
                |: :|  :  :       ::         ::: : |:|     |  :  | |::|    :
Db        740 TTTRNMTTSTNMTNNNNNNMWACTNNNNNNMWKAYYAHATNNWGCWWNNTDARRTNNTTV 681

Qy       2037 AGGCAACAGGTCTCATGTGGTCAACACTGGGTGGGCTATGTTGGCTCTCATTGATGCTGG 2096
              ::: ::   :| :: |  : ::                    : |||| : | :: ::|:
Db        680 MRRRWMTNTKTRWYSTTRRHHYTGATNNNNNNNNNNNNNNNNSCCTCTRMMTMRWTMKGD 621

Qy       2097 ACAGGCTAAGAGAGATTCGCAACCACTGCAC-CGGGCAGCTGCATACTTGATAAATTCCC 2155
              : :: ::  ::   |:::::| :::: :|  :: : : :|:  | |:  |:  ::::
Db        620 GMTVRKKVKWRDTTCTYVDVWADSWVWWYANWMRCRDVTYTRNNTYCKSYAHSYWYWSNN 561

Qy       2156 AATTGGAGGACGGTGACTTTCCG---CAGCAGGAAATAATGGGAGTCTTCAACAAGAATT 2212
              |::::: : :  :  | :||      : ::: ::| |::::      |: : :::  ::
Db        560 AMWYRRYSARNWSSMARWTTRNNWWMSGBVRMRWAGTMWWRHWNNNNTDTRYYWWWKRWA 501

Qy       2213 GCATGATCACATACGCCGCATACAGAAACATATTCCCCATTTGGGCGTTGGGAGAATACC 2272
              :: | ::  :  | :: :  : |  :: |   :   ::::|       |   ::|::::
Db        500 RBTTTVYDSMCNAKSMWRGNNWRAMKMWWAANNDAGAMDHWTYWMGNNTMWMRRAWKMMN 441

Qy       2273 AATCCCAAGTATTGCAATCTCGTTAATCGAGCCTTAGTTAGGGTGTCATCCTAATTATTT 2332
              :|:|  |        :| :::::::   ::::::|      ::  :: : :::: ::||
Db        440 MAWCRRAYCCNNNNNRACVWHKHKMWRWTWKYMWKAACNNNNBKAMYMRVAWMMYSRDTT 381

Qy       2333 CGACCTGTGGACAGTAAAAGAAATAATAATAATAACCTATTTTTCTTTTATTTCCATGGC 2392
                : :  |::::: :  ::  :: |:       :::|: | : :  ::  |:  :
Db        380 NTDWMMWTSDWBWHWYTVDYTMMRAWNNNNNNNNWRBCKTTSWMWWMDHMNTHCTYGNNTW 321

Qy       2393 TCTCTTAAAATGTTTGTGACTAATGAGTTTAGTAGTCAGCTAAAAAAAAAAGCAAACACG 2452
              : :: ||:          :|  : :: |::|:  |        |:  :   : :   ::
Db        320 GSAYBMAAMSMWAAGASNBVTYNWCWRMTYMGKTMTNNNNNNKAWYYRTKTVAWCNNRYY 261

Qy       2453 TGGAGAATGCCTGTAAGTTTTTTCTATTACTCATAGACGCCTCTCTTGCTTTCCCTGCAG 2512
              :: |:: ::  :::   ::::::::: :::::::::|:  ::   : :|  ::::  :
Db        260 YDTAVWTBKRNYKYCYAYBWYYBMYMGKHHWBWWRRABHRSWNMWWVKCRNKYMVSWHYH 201

Qy       2513 CAGAGGAATTAAATATATACATAAATAGAGATATAAAAAAAAAAA 2557
              :: :::|: :     ::: : ::::      : ::: | :::::
Db        200 AMRYBKWABAVGCNNNWKDRMAHHHWCATNNNMMWWWYAYMHHMH 156


RESULT 12
US-08-916-421B-1/c
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
;  APPLICANT: Bult et al.
;  TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanoco
; Patent No. 6503729
;  TITLE OF INVENTION:  jannaschii
;  FILE REFERENCE: PB275
;  CURRENT APPLICATION NUMBER: US/08/916,421B
;  CURRENT FILING DATE:  1997-08-22
;  PRIOR APPLICATION NUMBER: US 60/024,428
;  PRIOR FILING DATE: 1996-08-22
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 1664976
;   TYPE: DNA
;   ORGANISM: Methanococcus jannaschii
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (28222)..(28222)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (28257)..(28258)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84773)..(84773)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84808)..(84808)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84812)..(84812)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98120)..(98120)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98159)..(98159)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
```

```
; GENERAL INFORMATION:
;  APPLICANT: Keith Weinstock et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
;  TITLE OF INVENTION:  FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 107196.132
;  CURRENT APPLICATION NUMBER: US/09/248,796A
;  CURRENT FILING DATE:  1999-02-12
;  PRIOR APPLICATION NUMBER: US 60/074,725
;  PRIOR FILING DATE: 1998-02-13
;  PRIOR APPLICATION NUMBER: US 60/096,409
;  PRIOR FILING DATE: 1998-08-13
;  NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 3811
;   LENGTH: 747
;   TYPE: DNA
;   ORGANISM: Candida albicans
US-09-248-796A-3811

  Query Match             2.9%;  Score 74.4;  DB 4;  Length 747;
  Best Local Similarity   51.6%;  Pred. No. 1.6e-12;
  Matches  199;  Conservative    0;  Mismatches  181;  Indels    6;  Gaps    1;

Qy        505 GATGGTGGGTGGGGTTTGCATATTGAAGGTCCAAGCACCATGTTTGGCTCTGTCTTGAGT 564
              ||||||||| |||||  | |||  ||  | |  |    ||    |||||  |    | |||
Db        121 GATGGTGGTTGGGGACTTCATTCTGTTGATAAATATACTTGTTTTGGGACAACCATGAAT 180

Qy        565 TATATTACTCTGAGATTGCTAGGTGAGGGGCCTAATGATGGACAAGGGGAAATGGAGAAG 624
              |||  |   |||  | ||  ||||   ||       |  | |  ||    |   |||  |||
Db        181 TATGTATGTCTCCGGTTATTAGGAATGG------AAAAGGATCATCCAGTTTTGGTTAAG 234

Qy        625 GCACGTGACTGGATTCTAGGGCATGGTGGTGCTACTTATATAACGTCATGGGGGAAGATG 684
              ||  |  |     | |   |   ||||||| | | | |   |    |||||  |||
Db        235 GCGAGAAAAACATTACATCGTTTGGGTGGTGCCATTAAGAATCCACATTGGGGTAAGGCT 294

Qy        685 TGGCTTTCAGTACTTGGAGTGTATGAATGGTCTGGAAATAATCCCCTGCCCCCTGAGATA 744
              |||||  ||  |  |     | |||||||||   ||    || ||    || || ||  |
Db        295 TGGCTATCTATTTTGAATTTATATGAATGGGAGGGTGTGAACCCAGCTCCACCAGAACTT 354

Qy        745 TGGCTCCTTCCATACATGCTTCCATTTCATCCAGGAAGGATGTGGTGTCACTGCCGGATG 804
              |||    | ||  |||  | | |||  |||||||||   |    ||||   ||      ||
Db        355 TGGAGATTACCGTACTGGTTACCAATTCATCCAGCGAAATGGTGGGTACATACTAGGGCT 414

Qy        805 GTCTATTTGCCGATGTCCTACTTATATGGCAAGAGGTTTGTTGGTCCAATCTCACCAACA 864
              ||||||||||  ||   ||    | ||  || ||  ||    ||    |    |||
Db        415 ATCTATTTGCCATTGGGATATACGTCTGCAAACAGAGTTCAATGTGAGCTTGATCCACTT 474

Qy        865 GTATTATCTTTGAGAAAAGAGCTTTA 890
              ||  |    | |||||  ||  ||||
Db        475 TTAAAAGAGATCAGAAATGAAATTTA 500


RESULT 9
US-08-998-416-229/c
; Sequence 229, Application US/08998416
; Patent No. 6239264
;  GENERAL INFORMATION:
;    APPLICANT:  Philippsen, Peter
;    APPLICANT:  Pohlmann, Rainer
;    APPLICANT:  Steiner, Sabine
;    APPLICANT:  Mohr, Christine
;    APPLICANT:  Wendland, Jurgen
;    APPLICANT:  Knechtle, Philipp
;    APPLICANT:  Rebischung, Corinne
;    TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  1152
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6239264artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  No. 6239264th Carolina
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/998,416
;      FILING DATE:  24-DEC-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 0016/97
;      FILING DATE:  31-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  PF/5-30306/A/CGC1976
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  229:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  539 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  PAG1211RP
US-08-998-416-229

  Query Match             2.8%;  Score 70.4;  DB 3;  Length 539;
  Best Local Similarity   57.1%;  Pred. No. 2.4e-11;
  Matches  128;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy        604 GGACAAGGGGAAATGGAGAAGGCACGTGACTGGATTCTAGGGCATGGTGGTGCTACTTAT 663
              | |||    |  |||  ||  |||  ||  |     |  ||  |    || |||||||  |
Db        273 GAACACATCGTGATGCAGCGGGCGCGGAAGACCCTGCTGCGCATGGGGGGTGCTATTGGC 214

Qy        664 ATAACGTCATGGGGGAAGATGTGGCTTTCAGTACTTGGAGTGTATGAATGGTCTGGAAAT 723
                  |||  |||||  ||  ||||||||||  ||  ||      |  ||  |  |||   ||
Db        213 TCGCCGTACTGGGGTAAAGTGTGGCTTTCCGTGCTCAATTTATACAAGTGGGAGGGCGTG 154

Qy        724 AATCCCCTGCCCCCTGAGATATGGCTCCTTCCATACATGCTTCCATTTCATCCAGGAAGG 783
              || ||   |||  || |||  |||||||  || ||  |||   | ||||  | |||  ||  |
Db        153 AACCCAGCGCCACCAGAGCTATGGCTGCTGCCGTACCAGTTTCCTTATCACCCGGCTCGC 94

Qy        784 ATGTGGTGTCACTGCCGGATGGTCTATTTGCCGATGTCCTACTT 827
                ||||   ||    ||     |||  ||||||||||  |||| |
Db         93 TGGTGGGTACATACGCGTGCTATCTTTTTGCCGATGAGCTACCT 50


RESULT 10
US-09-313-294A-804
; Sequence 804, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
;  APPLICANT:  Lalgudi, Raghunath V.
;  APPLICANT:  Ito, Laura Y.
;  APPLICANT:  Sherman, Bradley K.
;  TITLE OF INVENTION:  POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
;  FILE REFERENCE:  PL-0017 US
;  CURRENT APPLICATION NUMBER: US/09/313,294A
;  CURRENT FILING DATE:  1999-05-14
;  NUMBER OF SEQ ID NOS:  7600
;  SOFTWARE:  PERL Program
; SEQ ID NO 804
;   LENGTH: 173
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. 6476212 700549863H1
US-09-313-294A-804
```

; APPLICANT: Karunanandaa, Balasulojini
; APPLICANT: Yu, Jaehyuk
; APPLICANT: Kishore, Ganesh M.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED
; TITLE OF INVENTION: WITH STEROL SYNTHESIS AND METABOLISM
; FILE REFERENCE: 16516.075
; CURRENT APPLICATION NUMBER: US/09/614,221A
; CURRENT FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/142,981
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 626
; SEQ ID NO 308
; LENGTH: 2196
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-614-221A-308

Query Match 8.2%; Score 211; DB 4; Length 2196;
Best Local Similarity 48.8%; Pred. No. 2.2e-55;
Matches 871; Conservative 0; Mismatches 860; Indels 54; Gaps 9;

```
Qy        505 GATGGTGGGTGGGGTTTGCATATTGAAGGTCCAAGCACCATGTTTGGCTCTGTCTTGAGT 564
Db        418 GATGGTGGCTGGGGTCTACATTCTGTTGACAAATCCACCGTGTTTGGTACAGTATTGAAC 477

Qy        565 TATATTACTCTGAGATTGCTAGGTGAGGGGCCTAATGATGGACAAGGGGAAATGGAGAAG 624
Db        478 TATGTAATCTTACGTTTATT------GGGTCTACCCAAGGACCACCCGGTTTGCGCCAAG 531

Qy        625 GCACGTGACTGGATTCTAGGGCATGGTGGTGCTACTTATATAACGTCATGGGGGAAGATG 684
Db        532 GCAAGAAGCACATTGTTAAGGTTAGGCGGTGCTATTGGATCCCCTCACTGGGGAAAAATT 591

Qy        685 TGGCTTTCAGTACTTGGAGTGTATGAATGGTCTGGAAATAATCCCCTGCCCCCTGAGATA 744
Db        592 TGGCTAAGTGCACTAAACTTGTATAAATGGGAAGGTGTGAACCCTGCCCCTCCTGAAACT 651

Qy        745 TGGCTCCTTCCATACATGCTTCCATTTCATCCAGGAAGGATGTGGTGTCACTGCCGGATG 804
Db        652 TGGTTACTTCCATATTCACTGCCCATGCATCCGGGGAGATGGTGGGTTCATACTAGAGGT 711

Qy        805 GTCTATTTGCCGATGTCCTACTTATATGGCAAGAGGTTTGTTGGTCCAATCTCACCAACA 864
Db        712 GTTTACATTCCGGTCAGTTACCTGTCATTGGTCAAATTTTCTTGCCCAATGACTCCTCTT 771

Qy        865 GTATTATCTTTGAGAAAAGAGCTTTATACAGTACCATACCATGATATAGATTGGGATCAG 924
Db        772 CTTGAAGAACTGAGGAATGAAATTTACACTAAACCGTTTGACAAGATTAACTTCTCCAAG 831

Qy        925 GCTCGCAATTTGTGTGCAAAGGAAGATTTGTACTATCCTCACCCACTTGTACAGGATATT 984
Db        832 AACAGGAATACCGTATGTGGAGTAGACCTATATTACCCCCATTCTACTACTTTGAATATT 891

Qy        985 CTTTGGGCATCTCTACACAAGTTCCTTGAGCCTATTCTGATGCATTGGCCTGGAAAAAGA 1044
Db        892 GCGAACAGCCTTGTAGTATTTTACGAAAAATACCTAAGAAACCGGTTCATTTACTCTCTA 951

Qy       1045 TTGAGGGAAAAGGCTATTATTTCTGCATTGGAGCATATACATTACGAAGATGAGAATACT 1104
Db        952 TCCAAGAAGAAGGTTTAT-----------GATCTAATCAAAACGGAGTTACAGAATACT 999

Qy       1105 CGATATATTTGCATAGGTCCTGTAAATAAGGTGTTAAATATGCTTTGCTGTTGGGTGGAA 1164
Db       1000 GATTCCTTGTGTATAGCACCTGTTAACCAGGCGTTTTGCGCACTTGTCACTCTTATTGAA 1059

Qy       1165 GATC---CAAATTCTGAGGCCTTCAAGTTGCATCTTCCCAGGATTTATGATTATCTATGG 1221
Db       1060 GAAGGGGTAGACTCGGAAGCGTTCCAGCGTCTCCAATATAGGTTCAAGGATGCATTGTTC 1119

Qy       1222 ATTGCAGAAGATGGCATGAAAATGCAGGGCTACAATGGAAGTCAACTATGGGACACTGCT 1281
Db       1120 CATGGTCCACAGGGTATGACCATTATGGGAACAAATGGTGTGCAAACCTGGGATTGTGCG 1179

Qy       1282 TTTGCTGTCCAAGCAATTATTGCATCTAACCTCATTGAAGAATTTGGTCCAACTATAAGA 1341
Db       1180 TTTGCCATTCAATACTTTTTCGTCGCAGGCCTCGCAGAAAGACCTGAATTCTATAACACA 1239

Qy       1342 AAAGCTCATACCTATATTAAGAATTCACAGGTTTTAGAAGATTGTCCAGGTGATCTTAAT 1401
Db       1240 ATTGTCTCTGCCTATAAATTCTTGTGTCATGCTCAATTTGACACCGAGTGCGTTCCAGGT 1299

Qy       1402 AAATGGTACCGTCACATTTCAAAAGGTGCTTGGCCTTTTTCAACTGGAGATCATGGATGG 1461
Db       1300 A---GTTATAGGGATAAGAGAAAGGGGGCTTGGGGCTTCTCAACAAAAACACAGGGCTAT 1356

Qy       1462 CCAATTTCTGACTGCACAGCTGAAGGACTGAAAGCTGTTCTATTACTATCCAAAATTGCA 1521
Db       1357 ACAGTGGCAGATTGCACTGCAGAAGCAATTAAAGCCATCATCATGGTGAAAAACTCTCCC 1416

Qy       1522 CCAGAAA---TAGTTGGTGAGCCAATAGACGTGAAGCGATTATATGATTCTGTAAATGTC 1578
Db       1417 GTCTTTAGTGAAGTACACCATATGATTAGCAGTGAACGTTTATTTGAAGGCATTGATGTG 1476

Qy       1579 ATTCTCTCACTACAGAA----------TGAAGATGGTGGTTTTGCAACATATGAGCTT 1626
Db       1477 TTATTGAACCTACAAAACATCGGATCTTTTGAATATGGTTCCTTTGCAACCTATGAAAAA 1536

Qy       1627 AAACGATCTTATAATTGGTTGGAGATAATCAATCCTGCTGAAACTTTTGGTGACATCGTT 1686
Db       1537 ATCAAGGCCCCACTAGCAATGGAAACCTTGAATCCTGCTGAAGTTTTTGGTGACATAATG 1596

Qy       1687 ATTGATTATCCTTATGTGGAATGTACATCAGCAGCGATTCAAGCTTTGGCATCATTTAGG 1746
Db       1597 GTAGAATACCCATACGTGGAATGTACTGATTCATCCGTTCTGGGGTTGACATATTTTCAC 1656

Qy       1747 AAATTATATCCTGGGCATCGCCGAGAAGAAATACAACATTGTATCGATAAAGCCACTACC 1806
Db       1657 AA---GTACTTCGACTATAGGAAAGAGGAAATACGTACACGCATCAGAATCGCCATCGAA 1713

Qy       1807 TTCATTGAAAAAATACAAGCTTCAGATGGATCATGGTATGGTTCTTGGGGAGTTTGCTTC 1866
Db       1714 TTCATAAAAAAATCTCAATTACCAGATGGAAGTTGGTATGGAAGCTGGGGTATTTGTTTT 1773

Qy       1867 ACTTACGGTGCTTGGTTTGGGGTAAAAGGTCTGATTGCTGCTGGAAGGAGTTTCAGTAAT 1926
Db       1774 ACATATGCCGGTATGTTTGCATTGGAGGCATTACACACCGTGGGGGAGACCTATGAGAAT 1833

Qy       1927 TGCTCAAGCATCCGTAAAGCTTGTGAATTTCTGCTGTCCAAGCAGCTTCCTTCTGGTGGC 1986
Db       1834 TCCTCAACGGTAAGAAAAGGTTGCGACTTCTTGGTCAGTAAACAGATGAAGGATGGCGGT 1893

Qy       1987 TGGGGAGAGAGTTATCTGTCCTGTCAAAACAAGGTGTATTCAAATCTGGAAGGCAACAGG 2046
Db       1894 TGGGGGGAATCAATGAAGTCCAGTGAA------TTACATAGTTATGTGGATAGTGAAAAA 1947

Qy       2047 TCTCATGTGGTCAACACTGGGTGGGCTATGTTGGCTCTCATTGATGCTGGACAGGCTAAG 2106
Db       1948 TCGCTAGTCGTTCAAACCGCATGGGCGCTAATTGCACTTCTTTTCGCTGAATATCCTAAT 2007

Qy       2107 AGAGATTCGCAACCACTGCACCGGGCAGCTGCATACTTGATAAATTCCCAATTGGAGGAC 2166
Db       2008 AAAG------AAGTCATCGACCGCGGTATTGACCTTTTAAAAAATAGACAAGAAGAATCC 2061

Qy       2167 GGTGACTTTCCGCAGCAGGAAATAATGGGAGTCTTCAACAAGAATTGCATGATCACATAC 2226
Db       2062 GGGGAATGGAAATTTGAAAGTGTAGAAGGTGTTTTCAACCACTCTTGTGCAATTGAATAC 2121

Qy       2227 GCCGCATACAGAAACATATTCCCCATTTGGGCGTTGGGAGAATAC 2271
Db       2122 CCAAGTTATCGATTCTTATTCCCTATTAAGGCATTAGGTATGTAC 2166
```

RESULT 8
US-09-248-796A-3811
; Sequence 3811, Application US/09248796A
; Patent No. 6747137

```
; GENERAL INFORMATION:
;  APPLICANT: WEI, Ming-Hui et al.
;  TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: CL001180
;  CURRENT APPLICATION NUMBER: US/09/815,048
;  CURRENT FILING DATE:  2001-03-23
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 4239
;   TYPE: DNA
;   ORGANISM: Human
US-09-815-048-1

  Query Match             10.9%;  Score 280.2;  DB 4;  Length 4239;
  Best Local Similarity   51.4%;  Pred. No. 4.5e-77;
  Matches  906;  Conservative    0;  Mismatches  798;  Indels   57;  Gaps    9;

Qy        522 GCATATTGAAGGTCCAAGCACCATGTTTGGCTCTGTCTTGAGTTATATTACTCTGAGATT 581
              ||| ||||| | |      |||| |||||||  |||    | | ||| |  |||| ||| |
Db        429 GCACATTGAGGATAAGTCCACCGTGTTTGGGACTGCGCTCAACTATGTGTCTCTCAGAAT 488

Qy        582 GCTAGGTGAGGGGCCTAATGATGGACAAGGGGAAATGGAGAAGGCACGTGACTGGATTCT 641
              || ||||  |||||| | |||          ||  |||     || ||  ||     |||
Db        489 TCTGGGTGTTGGGCCTGACGATCCT------GACCTGGTACGAGCCCGGAACATTCTTCA 542

Qy        642 AGGGCATGGTGGTGCTACTTATATAACGTCATGGGGGAAGATGTGGCTTTCAGTACTTGG 701
              | | |||||||||         ||  | || |||||||||| | |||||  | || ||
Db        543 CAAGAAAGGTGGTGCTGTGGCCATCCCCTCCTGGGGGAAGTTCTGGCTGGCTGTCCTGAA 602

Qy        702 AGTGTATGAATGGTCTGGAAATAATCCCCTGCCCCCTGAGATATGGCTCCTTCCATACAT 761
              || ||   |||  ||   ||| |||||  ||| |||||  |||||   ||||  ||
Db        603 TGTTTACAGCTGGGAAGGCCTCAATACCCTGTTCCCAGAGATGTGGCTGTTTCCTGACTG 662

Qy        762 GCTTCCATTTCATCCAGGAAGGATGTGGTGTCACTGCCGGATGGTCTATTTGCCGATGTC 821
              |   ||    || ||    |  | |||||  |||||||||  |||  ||  ||||  |||
Db        663 GGCACCGGCACACCCCTCCACACTCTGGTGCCACTGCCGGCAGGTGTACCTGCCCATGAG 722

Qy        822 CTACTTATATGGCAAGAGGTTTGTTGGTCCAATCTCACCAACAGTATTATCTTTGAGAAA 881
              |||||  || | |   || |    ||  |        ||   ||        |  |  |
Db        723 CTACTGCTACGCCGTTCGGCTGAGTGCCGCGGAAGACCCGCTGGTCCAGAGCCTCCGCCA 782

Qy        882 AGAGCTTTATACAGTACCATACCATGATATAGATTGGGATCAGGCTCGCAATTTGTGTGC 941
              ||||| |||   |   | |    || || |||      |   | ||          ||
Db        783 GGAGCTCTATGTGGAGGACTTCGCCAGCATTGACTGGCTGGCGCAGAGGAACAACGTGGC 842

Qy        942 AAAGGAAGATTTGTACTATCCTCACCCACTTGTACAGGATATTCTTTGGGCATCTCTACA 1001
              ||  ||  |||||.  || |||       | |      |  | |  ||   ||  |
Db        843 CCCCGACGAGCTGTACACGCCGCACAGCTGGCTGCTCCGCGTGGTATATGCGCTCCTCAA 902

Qy       1002 CAAGTTCCTTGAGCCTATTCTGATGCATTGGCCTGGAAAAAGATTGAGGGAAAAGGCTAT 1061
              |  ||                  |||    |          || ||  |   |||  |
Db        903 CCTGTA--------------TGAGCACCACCACAGTGCCCACCTGCGGCAGCGGGCCGT 947

Qy       1062 TATTTCTGCATTGGAGCATATACATTACGAAGATGAGAATACTCGATATATTTGCATAGG 1121
                  |  || || ||      |||  ||          ||      || |||| ||
Db        948 GCAGAAGCTGTATGAACACATTGTGGCCGACGACCGATTCACCAAGAGCATCAGCATCGG 1007

Qy       1122 TCCTGTAAATAAGGTGTTAAATATGCTT---TGCTGTTGGGTGGAAGATCCAAATTCTGA 1178
              ||  |    ||     | || ||||||     |||| |  |||||  |  ||    ||
Db       1008 CCCGATCTCGAAAACCATCAACATGCTTGTGCGCTGGTATGTGGACGGGCCCGCCTCCAC 1067

Qy       1179 GGCCTTCAAGTTGCATCTTCCCAGGATTTATGATTATCTATGGATTGCAGAAGATGGCAT 1238
              ||||||  ||  |||| |  |||| ||     || |||||  ||||  |      || |||||
Db       1068 TGCCTTCCAGGAGCATGTCTCCAGAATCCCGGACTATCTCTGGATGGGCCTTGACGGCAT 1127

Qy       1239 GAAAATGCAGGGCTACAATGGAAGTCAACTATGGGACACTGCTTTTGCTGTCCAAGCAAT 1298
              ||||||||||||  ||| ||    ||  | ||||||| || || ||  |||| ||  |
Db       1128 GAAAATGCAGGGCACCAACGGCTCACAGATCTGGGACACCGCATTCGCCATCCAGGCTCT 1187

Qy       1299 TATTGCATC------TAACCTCATTGAAGAATTTGGTCCAACTATAAGAAAAGCTCATAC 1352
```

```
                 |||   |        |||  ||     || |||    |      |    || ||||||
Db       1188 GCTTGAGGCGGGCGGGCACCACAGGCCCGAGTTTTCGTCCTGCCTGCAGAAGGCTCATGA 1247

Qy       1353 CTATATTAAGAATTCACAGGTTTTAGAAGATTGTCCAGGTGATCTTAATAAATGGTACCG 1412
              |   | | |  |||||||||   |||  |  |||  |   |  |  | ||| |  |||||
Db       1248 GTTCCTGAGGCTCTCACAGGTCCCAGATAACCCTCCCGACTACC---AGAAGTACTACCG 1304

Qy       1413 TCACATTTCAAAAGGTGCTTGGCCTTTTTCAACTGGAGATCATGGATGGCCAATTTCTGA 1472
              || ||    || ||||  |   | ||    ||    ||    || |||   |||||||
Db       1305 CCAGATGCGCAAGGGTGGCTTCTCCTTCAGTACGCTGGACTGCGGCTGGATCGTTTCTGA 1364

Qy       1473 CTGCACAGCTGAAGGACTGAAAGCTGTTCTATTACTATCCAAAATTGCACCAGAAATAGT 1532
              |||||| ||||| |   |||| ||||| ||  | ||     | |     ||  |    ||
Db       1365 CTGCACGGCTGAGGCCTTGAAGGCTGTGCTGCTCCTGCAGGAGAAGTGTCCCCA---TGT 1421

Qy       1533 TGGTGAGCCAATAGACGTGAAGCGATTATATGATTCTGTAAATGTCATTCTCTCACTACA 1592
                 ||||  ||     |  || | |  ||| |||   |||  | ||      |
Db       1422 CACCGAGCACATCCCCAGAGAACGGCTCTGCGATGCTGTGGCTGTGCTGCTGAACATGAG 1481

Qy       1593 GAATGAAGATGGTGGTTTTGCAACATATGAGCTTAAACGATCTTATAATTGGTTGGAGAT 1652
              |||  |||||| || || || || ||||||   || ||        | | | ||||| |
Db       1482 AAATCCAGATGGAGGGTTCGCCACCTATGAGACCAAGCGTGGGGGGCACTTGCTGGAGCT 1541

Qy       1653 AATCAATCCTGCTGAAACTTTTGGTGACATCGTTATTGATTATCCTTATGTGGAATGTAC 1712
              | || ||  | ||    || || |||||| | ||||| ||  | |||||||| || ||
Db       1542 GCTGAACCCCTCGGAGGTCTTCGGGGACATCATGATTGACTACACCTATGTGGAGTGCAC 1601

Qy       1713 ATCAGCAGCGATTCAAGCTTTGGCATCATTTAGGAAATTATATCCTGGGCATCGCCGAGA 1772
              |||||  | |||  || ||  |  ||  ||  |  || | ||| |  ||  |
Db       1602 CTCAGCCGTGATGCAGGCGCTTAAGTATTTCCACAAGCGTTTCCCGGAGCACAGGGCAGC 1661

Qy       1773 AGAAATACAACATTGTATCGATAAAGCCACTACCTTCATTGAAAAAATACAAGCTTCAGA 1832
              || || |   |    ||  | | |   ||| |    ||    | || | ||
Db       1662 GGAGATCCGGGAGACCCTCACGCAGGGCTTAGAGTTCTGTCGGCGGCAGCAGAGGGCCGA 1721

Qy       1833 TGGATCATGGTATGGTTCTTGGGGAGTTTGCTTCACTTACGGTGCTTGGTTTGGGGTAAA 1892
              ||| || ||| | || || |||||||||||||||||| |||||  | ||||||||  |  |
Db       1722 TGGCTCCTGGGAAGGCTCCTGGGGAGTTTGCTTCACCTACGGCACCTGGTTTGGCCTGGA 1781

Qy       1893 AGGTCTGATTGCTGCTGGAAGGAGTTTCAGTAAT---------TGCTCAAGCATCCGTAA 1943
              | |  |    |  |  || | | | ||  || |  |  ||    ||  ||  ||
Db       1782 GGCCTTCGCCTGTATGGGGCAGACCTACCGAGATGGGACTGCCTGTGCAGAGGTCTCCCG 1841

Qy       1944 AGCTTGTGAATTTCTGCTGTCCAAGCAGCTTCCTTCTGGTGGCTGGGGAGAGAGTTATCT 2003
              || ||||| || |||||||||  |||| |  |  || |||||||| |||  | |
Db       1842 GGCCTGTGACTTCCTGCTGTCCCGGCAGATGGCAGACGGAGGCTGGGGGGAGGACTTTGA 1901

Qy       2004 GTCCTGTCAAAACAAGGTGTATTCAAATCTGGAAGGCAACAGGTCTCATGTGGTCAACAC 2063
              ||||||      | ||| |      || || |  |  | ||| ||  |  ||||||
Db       1902 GTCCTG------CGAGGAGCGGCGTTATGTGCAGAGTGCCCAGTCCCAGATCCACAACAC 1955

Qy       2064 TGGGTGGGCTATGTTGGCTCTCATTGATGCTGGACAGGCTAAGAGAGATTCGCAACCACT 2123
                | ||||| ||| ||| || || | | | || | | | || |  | |||
Db       1956 ATGCTGGGCCATGATGGGGCTGATGGCCGTTCGGCATCCTGACATCGAGGCCCA------ 2009

Qy       2124 GCACCGGGCAGCTGCATACTTGATAAATTCCCAATTGGAGGACGGTGACTTTCCGCAGCA 2183
              | | | | ||      | | | |  || |  |  | || ||||  |||||| |
Db       2010 GGAGAGAGGAGTCCGGTGTCTACTTGAGAAACAGCTCCCCAATGGCGACTGGCCGCAGGA 2069

Qy       2184 GGAAATAATGGGAGTCTTCAACAAGAATTGCATGATCACATACGCCGCATACAGAAACAT 2243
              | ||    || ||||||||||||   ||    ||| | ||| |    ||||| ||||||
Db       2070 AAACATTGCTGGGGTCTTCAACAAGTCCTGTGCCATCTCCTACACGAGCTACAGGAACAT 2129

Qy       2244 ATTCCCCATTTGGGCGTTGGG 2264
              |||||||| |||||  | ||
Db       2130 CTTCCCCATCTGGGCCCTCGG 2150

RESULT 7
US-09-614-221A-308
; Sequence 308, Application US/09614221A
; Patent No. 6723837
; GENERAL INFORMATION:
```

```
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. 6727066 225080.16
US-09-919-039-112

  Query Match             11.9%;  Score 304.8;  DB 4;  Length 4346;
  Best Local Similarity   51.4%;  Pred. No. 7.6e-85;
  Matches  997;  Conservative    0;  Mismatches  877;  Indels   66;  Gaps   10;

Qy        343 CTCCAGTGCCATGACGGACACTGGCCGGGAGATTATGGAGGTCCCATGTTTCTGATGCCT 402
              || |||     | || || |||||| |||| |||||||| || ||  | || ||  ||||
Db        417 CTGCAGGCTGAGGATGGGCACTGGACGGGTGATTATGGTGGCCCACTTTTCCTCCTGCCA 476

Qy        403 GGCTTGGTAATTACTCTGTCTATCACTGGGGCGTTGAATACAGTCTTAACTGAAGAACAT 462
              ||| |  | || |||        |  |  |           |       | | |  | | |
Db        477 GGCCTCCTGATCACTTGCCACGTGGCACGC---------ATCCCTCTGCCAGCCGGATAC 527

Qy        463 AGAAAGGAAATATGCCGTTACCTCTATAATCATCAAAACAAGGATGGTGGGTGGGGTTTG 522
              ||| | || ||    || |||||          ||   |  || ||||| |||||  ||
Db        528 AGAGAAGAGATTGTGCGGTACCTGCGGTCAGTGCAGCTCCCTGACGGTGGCTGGGGCCTG 587

Qy        523 CATATTGAAGGTCCAAGCACCATGTTTGGCTCTGTCTTGAGTTATATTACTCTGAGATTG 582
              || ||||| | |     |||| |||||||  |||   | |  ||| |  |||| ||| | |
Db        588 CACATTGAGGATAAGTCCACCGTGTTTGGGACTGCGCTCAACTATGTGTCTCTCAGAATT 647

Qy        583 CTAGGTGAGGGGCCTAATGATGGACAAGGGGAAATGGAGAAGGCACGTGACTGGATTCTA 642
              || ||||  |||||| | |||          ||  |||     || ||  ||    |||
Db        648 CTGGGTGTTGGGCCTGACGAT------CCTGACCTGGTACGAGCCCGGAACATTCTTCAC 701

Qy        643 GGGCATGGTGGTGCTACTTATATAACGTCATGGGGGAAGATGTGGCTTTCAGTACTTGGA 702
                | | |||||||||      ||  | || |||||||||  | |||||  | || ||
Db        702 AAGAAAGGTGGTGCTGTGGCCATCCCCTCCTGGGGGAAGTTCTGGCTGGCTGTCCTGAAT 761

Qy        703 GTGTATGAATGGTCTGGAAATAATCCCCTGCCCCCTGAGATATGGCTCCTTCCATACATG 762
              || ||    |||   ||   ||| |||||  ||||| |||||  |||||  ||||  ||  |
Db        762 GTTTACAGCTGGGAAGGCCTCAATACCCTGTTCCCAGAGATGTGGCTGTTTCCTGACTGG 821

Qy        763 CTTCCATTTCATCCAGGAAGGATGTGGTGTCACTGCCGGATGGTCTATTTGCCGATGTCC 822
                ||    || ||     |   | ||||| |||||||||  ||| ||  |||| |||  |
Db        822 GCACCGGCACACCCCTCCACACTCTGGTGCCACTGCCGGCAGGTGTACCTGCCCATGAGC 881

Qy        823 TACTTATATGGCAAGAGGTTTGTTGGTCCAATCTCACCAACAGTATTATCTTTGAGAAAA 882
              ||||  || | |   || |   ||  |        ||     ||        |  |  |
Db        882 TACTGCTACGCCGTTCGGCTGAGTGCCGCGGAAGACCCGCTGGTCCAGAGCCTCCGCCAG 941

Qy        883 GAGCTTTATACAGTACCATACCATGATATAGATTGGGATCAGGCTCGCAATTTGTGTGCA 942
              ||||| |||    |     | |       || || |||      |    | ||       ||
Db        942 GAGCTCTATGTGGAGGACTTCGCCAGCATTGACTGGCTGGCGCAGAGGAACAACGTGGCC 1001

Qy        943 AAGGAAGATTTGTACTATCCTCACCCACTTGTACAGGATATTCTTTGGGCATCTCTACAC 1002
                || ||  |||||   || |||       | |     | | |  ||    ||  ||
Db       1002 CCCGACGAGCTGTACACGCCCCACAGCTGGCTGCTCCGCGTGGTATATGCGCTCCTCAAC 1061

Qy       1003 AAGTTCCTTGAGCCTATTCTGATGCATTGGCCTGGAAAAAGATTGAGGGAAAAGGCTATT 1062
                ||             |||    |   |       || || |   |||  |
Db       1062 CTGTA---------------TGAGCACCACCACAGTGCCCACCTGCGGCAGCGGGCCGTG 1106

Qy       1063 ATTTCTGCATTGGAGCATATACATTACGAAGATGAGAATACTCGATATATTTGCATAGGT 1122
                    |  || || ||       ||  ||          ||       || |||| ||
Db       1107 CAGAAGCTGTATGAACACATTGTGGCCGACGACCGATTCACCAAGAGCATCAGCATCGGC 1166

Qy       1123 CCTGTAAATAAGGTGTTAAATATGCTT---TGCTGTTGGGTGGAAGATCCAAATTCTGAG 1179
              ||  |    ||    | || ||||||     |||| |  ||||| |   ||     ||
Db       1167 CCGATCTCGAAAACCATCAACATGCTTGTGCGCTGGTATGTGGACGGGCCCGCCTCCACT 1226

Qy       1180 GCCTTCAAGTTGCATCTTCCCAGGATTTATGATTATCTATGGATTGCAGAAGATGGCATG 1239
              |||||| ||  |||| |  |||| ||    || |||||  |||||  |     || ||||||
Db       1227 GCCTTCCAGGAGCATGTCTCCAGAATCCCGGACTATCTCTGGATGGGCCTTGACGGCATG 1286

Qy       1240 AAAATGCAGGGCTACAATGGAAGTCAACTATGGGACACTGCTTTTGCTGTCCAAGCAATT 1299
              ||||||||||||  ||| ||    ||  | ||||||||  || || ||  |||| ||  |
Db       1287 AAAATGCAGGGCACCAACGGCTCACAGATCTGGGACACCGCATTCGCCATCCAGGCTCTG 1346

Qy       1300 ATTGCATC------TAACCTCATTGAAGAATTTGGTCCAACTATAAGAAAAGCTCATACC 1353
              |||  |         |||  ||     || |||    |     |    || ||||||
Db       1347 CTTGAGGCGGGCGGGCACCACAGGCCCGAGTTTTCGTCCTGCCTGCAGAAGGCTCATGAG 1406

Qy       1354 TATATTAAGAATTCACAGGTTTTAGAAGATTGTCCAGGTGATCTTAATAAATGGTACCGT 1413
              |   | | |  |||||||||   |||  |  |||  |  | |  | ||| |  |||||
Db       1407 TTCCTGAGGCTCTCACAGGTCCCAGATAACCCTCCCGACTACC---AGAAGTACTACCGC 1463

Qy       1414 CACATTTCAAAAGGTGCTTGGCCTTTTTCAACTGGAGATCATGGATGGCCAATTTCTGAC 1473
              || ||    || ||||  |    | ||     ||     ||    || |||    ||||||||
Db       1464 CAGATGCGCAAGGGTGGCTTCTCCTTCAGTACGCTGGACTGCGGCTGGATCGTTTCTGAC 1523

Qy       1474 TGCACAGCTGAAGGACTGAAAGCTGTTCTATTACTATCCAAAATTGCACCAGAAATAGTT 1533
              ||||| ||||| |    |||| ||||| ||  | ||     | |     ||  |    ||
Db       1524 TGCACGGCTGAGGCCTTGAAGGCTGTGCTGCTCCTGCAGGAGAAGTGTCCCCAT---GTC 1580

Qy       1534 GGTGAGCCAATAGACGTGAAGCGATTATATGATTCTGTAAATGTCATTCTCTCACTACAG 1593
                ||||  ||   |  | ||  | |  ||| ||||  |||  | ||      |
Db       1581 ACCGAGCACATCCCCAGAGAACGGCTCTGCGATGCTGTGGCTGTGCTGCTGAACATGAGA 1640

Qy       1594 AATGAAGATGGTGGTTTTGCAACATATGAGCTTAAACGATCTTATAATTGGTTGGAGATA 1653
              |||  |||||| ||  || || ||  ||||||   || ||          | | | ||||| |
Db       1641 AATCCAGATGGAGGGTTCGCCACCTATGAGACCAAGCGTGGGGGGCACTTGCTGGAGCTG 1700

Qy       1654 ATCAATCCTGCTGAAACTTTTGGTGACATCGTTATTGATTATCCTTATGTGGAATGTACA 1713
              | || ||  | ||    || || |||||  | ||||| ||  | ||||||| || ||
Db       1701 CTGAACCCCTCGGAGGTCTTCGGGGACATCATGATTGACTACACCTATGTGGAGTGCACC 1760

Qy       1714 TCAGCAGCGATTCAAGCTTTGGCATCATTTAGGAAATTATATCCTGGGCATCGCCGAGAA 1773
              ||||| | ||| || ||  |    | ||    ||     | || | |||  |   ||
Db       1761 TCAGCCGTGATGCAGGCGCTTAAGTATTTCCACAAGCGTTTCCCGGAGCACAGGGCAGCG 1820

Qy       1774 GAAATACAACATTGTATCGATAAAGCCACTACCTTCATTGAAAAAATACAAGCTTCAGAT 1833
              || || |  |    ||   | | |  |||  |    |                || |||
Db       1821 GAGATCCGGGAGACCCTCACGCAGGGCTTAGAGTTCTGTCGGCGGCAGCAGAGGGCCGAT 1880

Qy       1834 GGATCATGGTATGGTTCTTGGGGAGTTTGCTTCACTTACGGTGCTTGGTTTGGGGTAAAA 1893
              || || ||| | || || |||||||||||||||||  |||||  | ||||||||  |  |
Db       1881 GGCTCCTGGGAAGGCTCCTGGGGAGTTTGCTTCACCTACGGCACCTGGTTTGGCCTGGAG 1940

Qy       1894 GGTCTGATTGCTGCTGGAAGGAGTTTCAGTAAT---------TGCTCAAGCATCCGTAAA 1944
              |   |       |  ||  ||  | | |  ||          ||  ||    ||
Db       1941 GCCTTCGCCTGTATGGGGCAGACCTACCGAGATGGGACTGCCTGTGCAGAGGTCTCCCGG 2000

Qy       1945 GCTTGTGAATTTCTGCTGTCCAAGCAGCTTCCTTCTGGTGGCTGGGGAGAGAGTTATCTG 2004
              || ||||| || |||||||||  |||| |  |   || |||||||| |||  | |  |
Db       2001 GCCTGTGACTTCCTGCTGTCCCGGCAGATGGCAGACGGAGGCTGGGGGGAGGACTTTGAG 2060

Qy       2005 TCCTGTCAAAACAAGGTGTATTCAAATCTGGAAGGCAACAGGTCTCATGTGGTCAACACT 2064
              |||||  |  |   |  ||||    || |  |   |  ||| ||  |    |||||
Db       2061 TCCTGCGAGGAGCGGCGTTATT------TGCAGAGTGCCCAGTCCCAGATCCATAACACA 2114

Qy       2065 GGGTGGGCTATGTTGGCTCTCATTGATGCTGGACAGGCTAAGAGAGATTCGCAACCACTG 2124
              | ||||| ||| ||| || || |  | | || || | |  ||  | ||      |
Db       2115 TGCTGGGCCATGATGGGGCTGATGGCCGTTCGGCATCCTGACATCGAGGCCCA------G 2168

Qy       2125 CACCGGGCAGCTGCATACTTGATAAATTCCCAATTGGAGGACGGTGACTTTCCGCAGCAG 2184
              |  | | ||    |  |  | |  ||  |     | || |||||  ||||| |
Db       2169 GAGAGAGGAGTCCGGTGTCTACTTGAGAAACAGCTCCCCAATGGCGACTGGCCGCAGGAA 2228

Qy       2185 GAAATAATGGGAGTCTTCAACAAGAATTGCATGATCACATACGCCGCATACAGAAACATA 2244
              | ||    || |||||||||||  ||    ||| | ||| |    ||||| |||||
Db       2229 AACATTGCTGGGGTCTTCAACAAGTCCTGTGCCATCTCCTACACGAGCTACAGGAACATC 2288

Qy       2245 TTCCCCATTTGGGCGTTGGG 2264
              |||||||| |||||  | ||
Db       2289 TTCCCCATCTGGGCCCTCGG 2308


RESULT 6
US-09-815-048-1
; Sequence 1, Application US/09815048
; Patent No. 6737261
```

```
Qy       1887 GGTAAAAGGTCTGATTGCTGCTGGAAGGAGTTTCAGTAATTGCTCAAGCATCCGTAAAGC 1946
Db       1214 TGTAAAGGGATTAGTTGCTGCTGGTAGGACATTCAAAAACAGTCCTGCAATCAGAAAGGC 1273

Qy       1947 TTGTGAATTTCTGCTGTCCAAGCAGCTTCCTTCTGGTGGCTGGGGAGAGAGTTATCTGTC 2006
Db       1274 ATGTGACTTTTTGTTGTCAAAAGAGCTTCCTTCTGGAGGCTGGGGAGAAAGCTATTTGTC 1333

Qy       2007 CTGTCAAAACAAGGTGTATTCAAATCTGGAAGGCAACAGGTCTCATGTGGTCAACACTGG 2066
Db       1334 ATCCCAAGATCAGGTTTATACCAATCTCGAAGGGAAGCGACCTCATGCGGTGAACACTGG 1393

Qy       2067 GTGGGCTATGTTGGCTCTCATTGATGCTGGACAGGCTAAGAGAGATTCGCAACCACTGCA 2126
Db       1394 TTGGGCCATGCTAGCCCTAATCGATGCAGGGCAGGCTGAGAGAGATCCAATTCCTTTGCA 1453

Qy       2127 CCGGGCAGCTGCATACTTGATAAATTCCCAATTGGAGGACGGTGACTTTCCGCAGCAGGA 2186
Db       1454 TCGAGCAGCGAAGGTTTTGATCAACTTACAATCGGAAGATGGTGAATTTCCCCAGCAAGA 1513

Qy       2187 AATAATGGGAGTCTTCAACAAGAATTGCATGATCACATACGCCGCATACAGAAACATATT 2246
Db       1514 GATCATTGGAGTCTTCAACAAAAACTGCATGATCAGCTACTCCGAGTATAGAAACATCTT 1573

Qy       2247 CCCCATTTGGGCGTTGGGAGAATACCAATCCCAAGTATTG 2286
Db       1574 CCCTATTTGGGCCCTTGGGGAGTACCGTCGCCGCGTCTTG 1613
```

RESULT 4
US-09-419-679-7
; Sequence 7, Application US/09419679
; Patent No. 6630617
; GENERAL INFORMATION:
; APPLICANT: Falco, S. Carl
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Hitz, William D.
; APPLICANT: Kinney, Anthony J.
; APPLICANT: Rafalski, Antoni
; APPLICANT: McGonigle, Brian
; APPLICANT: Lohman, Karin
; TITLE OF INVENTION: Enzymes Involved In Squalene Metabolizm
; FILE REFERENCE: BB1113 US NA
; CURRENT APPLICATION NUMBER: US/09/419,679
; CURRENT FILING DATE: 1999-10-15
; EARLIER APPLICATION NUMBER: 60/105,405
; EARLIER FILING DATE: 1998-10-23
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 7
; LENGTH: 1300
; TYPE: DNA
; ORGANISM: Triticum aestivum
US-09-419-679-7

Query Match 22.0%; Score 563.8; DB 4; Length 1300;
Best Local Similarity 72.2%; Pred. No. 3.8e-167;
Matches 733; Conservative 0; Mismatches 282; Indels 0; Gaps 0;

```
Qy       1272 GGACACTGCTTTTGCTGTCCAAGCAATTATTGCATCTAACCTCATTGAAGAATTTGGTCC 1331
Db          7 GGACACAGCTTTTGCAGTTCAAGCTATTGCGGCCACTGACCTCATTGAAGAGTTTGCTCC 66

Qy       1332 AACTATAAGAAAAGCTCATACCTATATTAAGAATTCACAGGTTTTAGAAGATTGTCCAGG 1391
Db         67 CACTCTTAAGCTGGCACATGATTTTATTAAGAACTCTCAGGTTGTTGATGACTGCCCTGG 126

Qy       1392 TGATCTTAATAAATGGTACCGTCACATTTCAAAAGGTGCTTGGCCTTTTTCAACTGGAGA 1451
Db        127 AGATCTGAGTTACTGGTACCGTCACATATCTAAAGGTGCATGGCCCTTTTCTACAGCTGA 186

Qy       1452 TCATGGATGGCCAATTTCTGACTGCACAGCTGAAGGACTGAAAGCTGTTCTATTACTATC 1511
Db        187 TCATGGTTGGCCTATATCAGATTGCACTGCAGAAGGACTAAAGGCCTCATTATTGCTATC 246

Qy       1512 CAAAATTGCACCAGAAATAGTTGGTGAGCCAATAGACGTGAAGCGATTATATGATTCTGT 1571
Db        247 AAAGATTTCTCCAGAAATTGTGGGCGAATCGGTGGAAGTTAACAGACTATATGATGCTGT 306

Qy       1572 AAATGTCATTCTCTCACTACAGAATGAAGATGGTGGTTTTGCAACATATGAGCTTAAACG 1631
Db        307 CAATTGTTTGATGTCTTGGATGAATGAAAATGGTGGCTTCGCAACATATGAACTCCAAAG 366

Qy       1632 ATCTTATAATTGGTTGGAGATAATCAATCCTGCTGAAACTTTTGGTGACATCGTTATTGA 1691
Db        367 GTTTTATGCCTGGCTTGAGCTTATCAACCCTGCCGAGACATTCGGAGATATTGTGATTGA 426

Qy       1692 TTATCCTTATGTGGAATGTACATCAGCAGCGATTCAAGCTTTGGCATCATTTAGGAAATT 1751
Db        427 TTACCCGTATGTGGAATGTACCTCAGCCGCAATTCAGGCCCTGACATCATTTAAAAAGCT 486

Qy       1752 ATATCCTGGGCATCGCCGAGAAGAAATACAACATTGTATCGATAAAGCCACTACCTTCAT 1811
Db        487 CTATCCTGGGCACCGCAGGAAAGATGTAGATAACTGTATCAACAAAGCTGCTAGTTACAT 546

Qy       1812 TGAAAAAATACAAGCTTCAGATGGATCATGGTATGGTTCTTGGGGAGTTTGCTTCACTTA 1871
Db        547 TGAGAGCATCCAAAGAAAAGATGGTTCATGGTATGGCTCTTGGGCTGTGTGCTTCACCTA 606

Qy       1872 CGGTGCTTGGTTTGGGGTAAAAGGTCTGATTGCTGCTGGAAGGAGTTTCAGTAATTGCTC 1931
Db        607 TGGCACATGGTTCGGAGTGAAGGGGCTACTAGCTGCAGGTAGAACCTTCAAGAGCAGTCC 666

Qy       1932 AAGCATCCGTAAAGCTTGTGAATTTCTGCTGTCCAAGCAGCTTCCTTCTGGTGGCTGGGG 1991
Db        667 TGCAATCAGAAAGGCATGTGACTTTCTGATGTCAAAAGAGCTTCCTTTCGGTGGCTGGGG 726

Qy       1992 AGAGAGTTATCTGTCCTGTCAAAACAAGGTGTATTCAAATCTGGAAGGCAACAGGTCTCA 2051
Db        727 AGAAAGCTATCTGTCATCTCAAGATCAGGTTTACACCAATCTTGAAGGGAAGCATACTCA 786

Qy       2052 TGTGGTCAACACTGGGTGGGCTATGTTGGCTCTCATTGATGCTGGACAGGCTAAGAGAGA 2111
Db        787 TGCTGTCAACACTGGCTGGGCCATGCTGACTCTAATTGACGCAGGACAGGCTGAGAGAGA 846

Qy       2112 TTCGCAACCACTGCACCGGGCAGCTGCATACTTGATAAATTCCCAATTGGAGGACGGTGA 2171
Db        847 CCCGACGCCTCTGCATCGAGCAGCGAAGGTTTTGATAAACTTACAATCAGAGGATGGGGA 906

Qy       2172 CTTTCCGCAGCAGGAAATAATGGGAGTCTTCAACAAGAATTGCATGATCACATACGCCGC 2231
Db        907 ATTTCCTCAGCAAGAGATCATGGGAGTCTTCAACAAGAACTGCATGATCAGCTACTCCCA 966

Qy       2232 ATACAGAAACATATTCCCCATTTGGGCGTTGGGAGAATACCAATCCCAAGTATTG 2286
Db        967 GTATCGGAACATCTTCCCTATCTGGGCGCTTGGCGAGTACCGCTGCCGGGTGCTG 1021
```

RESULT 5
US-09-919-039-112
; Sequence 112, Application US/09919039
; Patent No. 6727066
; GENERAL INFORMATION:
; APPLICANT: Kaser, Matthew R.
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
; FILE REFERENCE: PA-0035 US
; CURRENT APPLICATION NUMBER: US/09/919,039
; CURRENT FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 60/222,113
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 401
; SOFTWARE: PERL Program
; SEQ ID NO 112
; LENGTH: 4346
; TYPE: DNA
; ORGANISM: Homo sapiens

```
Db       1623 AATTTCATCGAGAGTATTCAGAAAAGCGATGGTTCATGGTATGGCTCTTGGGCCGTCTGT 1682

Qy       1864 TTCACTTACGGTGCTTGGTTTGGGGTAAAAGGTCTGATTGCTGCTGGAAGGAGTTTCAGT 1923

Db       1683 TTCACATACGGCACTTGGTTTGGTGTGAAGGGACTAATTGCTGCTGGTAGAACATTTGAG 1742

Qy       1924 AATTGCTCAAGCATCCGTAAAGCTTGTGAATTTCTGCTGTCCAAGCAGCTTCCTTCTGGT 1983

Db       1743 AACAGTCCTGCAATTAGAAAGGCATGCGACTTTCTGTTGTCAAAAGAACTTCCTTCCGGT 1802

Qy       1984 GGCTGGGGAGAGAGTTATCTGTCCTGTCAAAACAAGGTGTATTCAAATCTGGAAGGCAAC 2043

Db       1803 GGTTGGGGAGAAAGCTATTTGTCATCTCAAGACCAGGTTTACACCAATCTCAAAGGCAAC 1862

Qy       2044 AGGTCTCATGTGGTCAACACTGGGTGGGCTATGTTGGCTCTCATTGATGCTGGACAGGCT 2103

Db       1863 CGGCCTCATGCGGTGAACACTAGTTGGGCCATGCTGGCGCTGATTGATGCGGGCCAGGCC 1922

Qy       2104 AAGAGAGATTCGCAACCACTGCACCGGGCAGCTGCATACTTGATAAATTCCCAATTGGAG 2163

Db       1923 GAGAGAGATCCAACGCCTCTACACCGAGCAGCAAAGGTTTTGATCAACTTACAATCAGAG 1982

Qy       2164 GACGGTGACTTTCCGCAGCAGGAAATAATGGGAGTCTTCAACAAGAATTGCATGATCACA 2223

Db       1983 GACGGAGAATTTCCTCAGCAAGAGATCATAGGAGTGTTCAACAAGAACTGCATGATAAGC 2042

Qy       2224 TACGCCGCATACAGAAACATATTCCCCATTTGGGCGTTGGGAGAATACCAATCCCAAGTA 2283

Db       2043 TACTCCCAGTACAGGAACATCTTCCCGATTTGGGCTCTGGGTGAGTACCGGTGTCGAGTC 2102

Qy       2284 TTGCAATCTCGTTAATCGAGCCTTAGTTAGGGTG 2317

Db       2103 TTGGGGGCTGGCAAGCCTTGGCGGTGAACGGGAG 2136


RESULT 3
US-09-419-679-3
; Sequence 3, Application US/09419679
; Patent No. 6630617
; GENERAL INFORMATION:
;  APPLICANT: Falco, S. Carl
;  APPLICANT: Famodu, Omolayo O.
;  APPLICANT: Hitz, William D.
;  APPLICANT: Kinney, Anthony J.
;  APPLICANT: Rafalski, Antoni
;  APPLICANT: McGonigle, Brian
;  APPLICANT: Lohman, Karin
;  TITLE OF INVENTION: Enzymes Involved In Squalene Metabolizm
;  FILE REFERENCE: BB1113 US NA
;  CURRENT APPLICATION NUMBER: US/09/419,679
;  CURRENT FILING DATE: 1999-10-15
;  EARLIER APPLICATION NUMBER: 60/105,405
;  EARLIER FILING DATE: 1998-10-23
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 3
;   LENGTH: 1882
;   TYPE: DNA
;   ORGANISM: Oryza sativa
US-09-419-679-3

  Query Match             36.9%;  Score 944;  DB 4;  Length 1882;
  Best Local Similarity   74.4%;  Pred. No. 2.7e-287;
  Matches 1190;  Conservative    0;  Mismatches  410;  Indels    0;  Gaps    0;

Qy        687 GCTTTCAGTACTTGGAGTGTATGAATGGTCTGGAAATAATCCCCTGCCCCCTGAGATATG 746

Db         14 GCTTTCGGTGCTTGGTGTATTTGACTGGTCTGGCAACAACCCAGTGCCACCAGAAATATG 73

Qy        747 GCTCCTTCCATACATGCTTCCATTTCATCCAGGAAGGATGTGGTGTCACTGCCGGATGGT 806

Db         74 GTTGTTGCCATATTTCCTGCCGATTCATCCAGGGCGAATGTGGTGTCATTGCCGGATGGT 133

Qy        807 CTATTTGCCGATGTCCTACTTATATGGCAAGAGGTTTGTTGGTCCAATCTCACCAACAGT 866

Db        134 TTATTTGCCTATGTGTTACATTTATGGAAAGAGGTTTGTGGGCCCAGTTACACCAATTAT 193

Qy        867 ATTATCTTTGAGAAAAGAGCTTTATACAGTACCATACCATGATATAGATTGGGATCAGGC 926

Db        194 ATTGGAATTAAGAAAGGAACTCTACGAAGTACCCTACAATGAAGTTGATTGGGACAAGGC 253

Qy        927 TCGCAATTTGTGTGCAAAGGAAGATTTGTACTATCCTCACCCACTTGTACAGGATATTCT 986

Db        254 TCGCAATCTATGTGCTAAGGAAGATCTGTACTATCCACATCCATTCGTGCAGGATGTATT 313

Qy        987 TTGGGCATCTCTACACAAGTTCCTTGAGCCTATTCTGATGCATTGGCCTGGAAAAAGATT 1046

Db        314 ATGGGCCACTCTCCACAAATTTGTTGAACCAGCTATGTTGCGTTGGCCTGGGAACAAATT 373

Qy       1047 GAGGGAAAAGGCTATTATTTCTGCATTGGAGCATATACATTACGAAGATGAGAATACTCG 1106

Db        374 GAGGGAGAAAGCTTTGGACACTGTCATGCAGCATATTCATTATGAAGATGAGAACACCCG 433

Qy       1107 ATATATTTGCATAGGTCCTGTAAATAAGGTGTTAAATATGCTTTGCTGTTGGGTGGAAGA 1166

Db        434 ATATATTTGCATTGGTCCAGTAAACAAGGTATTAAATATGCTTGCTTGCTGGATTGAAGA 493

Qy       1167 TCCAAATTCTGAGGCCTTCAAGTTGCATCTTCCCAGGATTTATGATTATCTATGGATTGC 1226

Db        494 TCCAAACTCAGAGGCATTCAAACTCCACATTCCAAGAGTCCACGATTACCTATGGATTGC 553

Qy       1227 AGAAGATGGCATGAAAATGCAGGGCTACAATGGAAGTCAACTATGGGACACTGCTTTTGC 1286

Db        554 AGAAGATGGCATGAAAATGCAGGGTTATAATGGAAGCCAGCTGTGGGACACAGCTTTCAC 613

Qy       1287 TGTCCAAGCAATTATTGCATCTAACCTCATTGAAGAATTTGGTCCAACTATAAGAAAAGC 1346

Db        614 AGTTCAAGCTATAGTGGCTACTGGCCTCATTGAAGAATTTGGTCCTACTCTTAAACTAGC 673

Qy       1347 TCATACCTATATTAAGAATTCACAGGTTTTAGAAGATTGTCCAGGTGATCTTAATAAATG 1406

Db        674 ACATGGCTACATAAAGAAAACGCAGGTTATCGATGACTGCCCTGGAGATCTTAGTCAGTG 733

Qy       1407 GTACCGTCACATTTCAAAAGGTGCTTGGCCTTTTTCAACTGGAGATCATGGATGGCCAAT 1466

Db        734 GTACCGCCACATATCTAAAGGTGCATGGCCCTTTTCTACTGCTGATCATGGTTGGCCTAT 793

Qy       1467 TTCTGACTGCACAGCTGAAGGACTGAAAGCTGTTCTATTACTATCCAAAATTGCACCAGA 1526

Db        794 ATCAGATTGCACTGCAGAAGGACTTAAGGCGGCATTATTGCTATCGAAGATTTCTCCAGA 853

Qy       1527 AATAGTTGGTGAGCCAATAGACGTGAAGCGATTATATGATTCTGTAAATGTCATTCTCTC 1586

Db        854 TATTGTTGGCGAAGCAGTGGAAGTTAATAGACTGTATGATTCTGTCAATTGTTTGATGTC 913

Qy       1587 ACTACAGAATGAAGATGGTGGTTTTGCAACATATGAGCTTAAACGATCTTATAATTGGTT 1646

Db        914 ATACATGAATGATAATGGTGGATTTGCAACATATGAACTCACAAGGTCTTATGCCTGGCT 973

Qy       1647 GGAGATAATCAATCCTGCTGAAACTTTTGGTGACATCGTTATTGATTATCCTTATGTGGA 1706

Db        974 GGAGCTTATCAATCCTGCTGAGACCTTTGGGGACATTGTGATTGATTATCCTTATGTGGA 1033

Qy       1707 ATGTACATCAGCAGCGATTCAAGCTTTGGCATCATTTAGGAAATTATATCCTGGGCATCG 1766

Db       1034 ATGCACTTCAGCAGCAATTCAGGCCCTGACAGCATTTAAAAAGCTCTACCCTGGACACCG 1093

Qy       1767 CCGAGAAGAAATACAACATTGTATCGATAAAGCCACTACCTTCATTGAAAAAATACAAGC 1826

Db       1094 CAAGAGTGAAATAGACAACTGTATAAGCAAAGCTGCTAGCTTTATTGAGGGTATTCAAAA 1153

Qy       1827 TTCAGATGGATCATGGTATGGTTCTTGGGGAGTTTGCTTCACTTACGGTGCTTGGTTTGG 1886

Db       1154 AAGCGATGGTTCATGGTATGGTTCTTGGGCTGTTTGTTTTACCTATGGCACATGGTTTGG 1213
```

```
Db       2461 GCCTGTAAGTTTTTTCTATTACTCATAGACGCCTCTCTTGCTTTCCCTGCAGCAGAGGAA 2520

Qy       2521 TTAAATATATACATAAATAGAGATATAAAAAAAAAAAAAA 2560
              |||||||||||||||||||||||||||||||||||||||||
Db       2521 TTAAATATATACATAAATAGAGATATAAAAAAAAAAAAAA 2560


RESULT 2
US-09-419-679-1
; Sequence 1, Application US/09419679
; Patent No. 6630617
; GENERAL INFORMATION:
;  APPLICANT: Falco, S. Carl
;  APPLICANT: Famodu, Omolayo O.
;  APPLICANT: Hitz, William D.
;  APPLICANT: Kinney, Anthony J.
;  APPLICANT: Rafalski, Antoni
;  APPLICANT: McGonigle, Brian
;  APPLICANT: Lohman, Karin
;  TITLE OF INVENTION: Enzymes Involved In Squalene Metabolizm
;  FILE REFERENCE: BB1113 US NA
;  CURRENT APPLICATION NUMBER: US/09/419,679
;  CURRENT FILING DATE: 1999-10-15
;  EARLIER APPLICATION NUMBER: 60/105,405
;  EARLIER FILING DATE: 1998-10-23
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 1
;   LENGTH: 2558
;   TYPE: DNA
;   ORGANISM: Zea mays
US-09-419-679-1

  Query Match             46.2%;  Score 1183.6;  DB 4;  Length 2558;
  Best Local Similarity   72.2%;  Pred. No. 0;
  Matches 1540;  Conservative    0;  Mismatches  594;  Indels    0;  Gaps    0;

Qy        184 ACCCACAAACACAGCGCTGATCTACTTATGCGGATGCAGTTCGCAAGAGAGAACCCAACA 243
              ||   ||| |||||||| || || || ||||| || |||||||| | ||| ||| | | |
Db          3 ACGAGCAAGCACAGCGCCGACCTCCTCATGCGCATCCAGTTCGCCAAAGAAAACTCGATT 62

Qy        244 CGTGAAGTCTTGCCCAAAGTCGGAGTTAAGGATATTGAGGATGTGACCCAAGAGATTGTG 303
                | | ||   ||   |   ||   ||| ||||||||| | ||   ||||
Db         63 GAGCTTCACCTTCCAGGCATCAAGCTCGGTGAGCATGAAGATGTGACCGAGGAAGCTGTG 122

Qy        304 ACAAAAACATTAAGAAGGGCCGTAAGTTTCCATTCAACTCTCCAGTGCCATGACGGACAC 363
                | ||||| | |||||| | ||        || |||||||||   |||||| ||||||
Db        123 TTGACTACATTGAAAAGGGCAATCAGCCGTTTCTCTACTCTCCAGGCACATGATGGACAC 182

Qy        364 TGGCCGGGAGATTATGGAGGTCCCATGTTTCTGATGCCTGGCTTGGTAATTACTCTGTCT 423
              ||||| || |||||||| ||||| ||||| || ||||| |||||| | || ||  ||| |
Db        183 TGGCCTGGGGATTATGGTGGTCCTATGTTCCTTATGCCAGGCTTGATCATAACATTGTAT 242

Qy        424 ATCACTGGGGCGTTGAATACAGTCTTAACTGAAGAACATAGAAAGGAAATATGCCGTTAC 483
                | |||||  ||  | || ||  |||||   |  |||||     |||||  ||  |||| ||
Db        243 GTGACTGGAGCACTAAACACTGTCTTGTCATTGGAACACCAGAAGGAGATCCGCCGGTAT 302

Qy        484 CTCTATAATCATCAAAACAAGGATGGTGGGTGGGGTTTGCATATTGAAGGTCCAAGCACC 543
              || |||||||| || ||  | ||||| || ||||| ||||| ||||| ||||||||||||
Db        303 CTTTATAATCACCAGAATGAAGATGGCGGCTGGGGCTTGCACATTGAGGGTCCAAGCACC 362

Qy        544 ATGTTTGGCTCTGTCTTGAGTTATATTACTCTGAGATTGCTAGGTGAGGGGCCTAATGAT 603
              ||||| ||||| |   |||  ||| ||| | |||||||||| || ||||| ||  ||  |
Db        363 ATGTTCGGCTCAGCACTGACCTATGTTATTTTGAGATTGCTTGGAGAGGGACCAGATAGT 422

Qy        604 GGACAAGGGGAAATGGAGAAGGCACGTGACTGGATTCTAGGGCATGGTGGTGCTACTTAT 663
              ||| | || |  |||||||| |  ||  |||||||  |||  |||||||| || || |||
Db        423 GGAGATGGAGCCATGGAGAAAGGTCGAAACTGGATATTAGACCATGGTGGAGCAACATAT 482

Qy        664 ATAACGTCATGGGGGAAGATGTGGCTTTCAGTACTTGGAGTGTATGAATGGTCTGGAAAT 723
              ||||| || |||||||||| | |||||||| ||||| || || | ||||||||||||| ||
Db        483 ATAACATCGTGGGGGAAGTTTTGGCTTTCGGTACTAGGTGTATTTGAATGGTCTGGTAAC 542

Qy        724 AATCCCCTGCCCCCTGAGATATGGCTCCTTCCATACATGCTTCCATTTCATCCAGGAAGG 783
              || ||  |||| || ||  ||||||| || |||||  | || |||||||| ||||| |||
Db        543 AACCCGGTGCCACCAGAAGTATGGCTACTGCCATATCTCCTCCCATTTCACCCAGGGAGG 602

Qy        784 ATGTGGTGTCACTGCCGGATGGTCTATTTGCCGATGTCCTACTTATATGGCAAGAGGTTT 843
              |||||||||| || || ||||| |||||||| |||| |||| | ||||| |||||||||
Db        603 ATGTGGTGTCATTGTCGAATGGTGTATTTGCCAATGTGCTACATTTATGGGAAGAGGTTT 662

Qy        844 GTTGGTCCAATCTCACCAACAGTATTATCTTTGAGAAAAGAGCTTTATACAGTACCATAC 903
              ||||| | |||| ||||| |  || || || || || |||| | ||  || |||
Db        663 GTTGGCCGAATCACACCACTTCTGTTGGAATTAAGGAAGGAACTTTTCAAAGACCCTTAC 722

Qy        904 CATGATATAGATTGGGATCAGGCTCGCAATTTGTGTGCAAAGGAAGATTTGTACTATCCT 963
                | || ||||||||  |||| |||||  | ||||| |||||||||| ||||||| ||
Db        723 AGCAAGATTGATTGGGACAAGGCCCGCAACCTATGTGCCAAGGAAGATCTGTACTACCCA 782

Qy        964 CACCCACTTGTACAGGATATTCTTTGGGCATCTCTACACAAGTTCCTTGAGCCTATTCTG 1023
              |||||| | || || ||| |  | |||||  |||| || || ||| |||| ||  || ||
Db        783 CACCCATTCGTTCAAGATGTGTTGTGGGCCACTCTCCATAAATTCGTTGAACCAGTTATG 842

Qy       1024 ATGCATTGGCCTGGAAAAAGATTGAGGGAAAAGGCTATTATTTCTGCATTGGAGCATATA 1083
              |||||||||||||| |  | ||||||||| || ||| |     | |   || | ||| |
Db        843 ATGCATTGGCCTGGCAGCAAATTGAGGGAGAAAGCTCTGGAAACAGTCATGCAACATGTT 902

Qy       1084 CATTACGAAGATGAGAATACTCGATATATTTGCATAGGTCCTGTAAATAAGGTGTTAAAT 1143
              ||||| |||||||||| ||||| |||||||||| |||||||||| ||||| || |||
Db        903 CATTATGAAGATGAGAACACTCGTTATATTTGCATTGGTCCTGTAAACAAGGTATTGAAT 962

Qy       1144 ATGCTTTGCTGTTGGGTGGAAGATCCAAATTCTGAGGCCTTCAAGTTGCATCTTCCCAGG 1203
              ||||||   || ||| | |||||||||| || ||||||||||  | ||| | ||  |
Db        963 ATGCTTGCTTGCTGGATTGAAGATCCAAACTCGGAGGCCTTCAAACTTCATATCCCACGA 1022

Qy       1204 ATTTATGATTATCTATGGATTGCAGAAGATGGCATGAAAATGCAGGGCTACAATGGAAGT 1263
                | |||||||  | ||| |||| |||||||||||||| |||||||| || ||||| ||
Db       1023 GTCTATGATTACTTGTGGCTTGCTGAAGATGGCATGAAGATGCAGGGTTATAATGGCAGC 1082

Qy       1264 CAACTATGGGACACTGCTTTTGCTGTCCAAGCAATTATTGCATCTAACCTCATTGAAGAA 1323
              |||||  ||||| || |||||  | || ||||| ||| | ||  | ||||| |||||||||
Db       1083 CAACTTTGGGATACAGCTTTCACAGTTCAAGCCATTGTGGCTACCAACCTTATTGAAGAG 1142

Qy       1324 TTTGGTCCAACTATAAGAAAAGCTCATACCTATATTAAGAATTCACAGGTTTTAGAAGAT 1383
              |||||||| ||  | | |  ||| || | |||||| |||||||||||||| | || ||
Db       1143 TTTGGTCCTACCCTTAAACTAGCACACAACTATATCAAGAATTCACAGGTTCTTGATGAC 1202

Qy       1384 TGTCCAGGTGATCTTAATAAATGGTACCGTCACATTTCAAAAGGTGCTTGGCCTTTTTCA 1443
              || || |||||||| ||| | |||||||| |||| || |||||||| ||||| || |||
Db       1203 TGCCCTGGTGATCTGAATGACTGGTACCGCCACACATCTAAAGGTGCATGGCCATTCTCA 1262

Qy       1444 ACTGGAGATCATGGATGGCCAATTTCTGACTGCACAGCTGAAGGACTGAAAGCTGTTCTA 1503
              |||| |||||||| ||||| || ||||| ||||| |||||||||| || |||    ||
Db       1263 ACTGCTGATCATGGTTGGCCTATATCTGATTGCACTGCTGAAGGACTAAAGGCTTCATTA 1322

Qy       1504 TTACTATCCAAAATTGCACCAGAAATAGTTGGTGAGCCAATAGACGTGAAGCGATTATAT 1563
              || |||| |  ||  | ||  |||| |||||||| || || ||  |  |||| |||
Db       1323 TTGTTATCAAGGATCTCTCCCAAAATTGTTGGTGAACCGATGGAAGCTAATAGATTTTAT 1382

Qy       1564 GATTCTGTAAATGTCATTCTCTCACTACAGAATGAAGATGGTGGTTTTGCAACATATGAG 1623
              ||| |||| | |    | | ||        ||||||  |||| ||||| || ||||||||
Db       1383 GATGCTGTCAGTTGTCTGATGTCTTATATGAATGATAATGGCGGTTTCGCGACATATGAA 1442

Qy       1624 CTTAAACGATCTTATAATTGGTTGGAGATAATCAATCCTGCTGAAACTTTTGGTGACATC 1683
              || | | ||||||||  |||||||| | |||||||| ||||| || |||| || ||
Db       1443 CTCACAAGATCTTATCCCTGGTTGGAGCTGATCAATCCCGCTGAGACCTTTGGGGATATT 1502

Qy       1684 GTTATTGATTATCCTTATGTGGAATGTACATCAGCAGCGATTCAAGCTTTGGCATCATTT 1743
              || |||||||| || ||||| |||||||||||||||||| ||||| ||  || |||||||
Db       1503 GTGATTGATTACCCGTATGTTGAATGTACATCAGCAGCAATTCAGGCCCTGACATCATTC 1562

Qy       1744 AGGAAATTATATCCTGGGCATCGCCGAGAAGAAATACAACATTGTATCGATAAAGCCACT 1803
              |  ||| |||| |||||||| ||| |  |||| |  |  | |||||||  |||||  |
Db       1563 AAAAAACTATACCCTGGGCACCGCAGGAAAGAGGTGGATAACTGTATCAGCAAAGCTTCC 1622

Qy       1804 ACCTTCATTGAAAAAATACAAGCTTCAGATGGATCATGGTATGGTTCTTGGGGAGTTTGC 1863
```

```
Qy       301 GTGACAAAAACATTAAGAAGGGCCGTAAGTTTCCATTCAACTCTCCAGTGCCATGACGGA 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 GTGACAAAAACATTAAGAAGGGCCGTAAGTTTCCATTCAACTCTCCAGTGCCATGACGGA 360

Qy       361 CACTGGCCGGGAGATTATGGAGGTCCCATGTTTCTGATGCCTGGCTTGGTAATTACTCTG 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 CACTGGCCGGGAGATTATGGAGGTCCCATGTTTCTGATGCCTGGCTTGGTAATTACTCTG 420

Qy       421 TCTATCACTGGGGCGTTGAATACAGTCTTAACTGAAGAACATAGAAAGGAAATATGCCGT 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 TCTATCACTGGGGCGTTGAATACAGTCTTAACTGAAGAACATAGAAAGGAAATATGCCGT 480

Qy       481 TACCTCTATAATCATCAAAACAAGGATGGTGGGTGGGGTTTGCATATTGAAGGTCCAAGC 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 TACCTCTATAATCATCAAAACAAGGATGGTGGGTGGGGTTTGCATATTGAAGGTCCAAGC 540

Qy       541 ACCATGTTTGGCTCTGTCTTGAGTTATATTACTCTGAGATTGCTAGGTGAGGGGCCTAAT 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 ACCATGTTTGGCTCTGTCTTGAGTTATATTACTCTGAGATTGCTAGGTGAGGGGCCTAAT 600

Qy       601 GATGGACAAGGGGAAATGGAGAAGGCACGTGACTGGATTCTAGGGCATGGTGGTGCTACT 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 GATGGACAAGGGGAAATGGAGAAGGCACGTGACTGGATTCTAGGGCATGGTGGTGCTACT 660

Qy       661 TATATAACGTCATGGGGGAAGATGTGGCTTTCAGTACTTGGAGTGTATGAATGGTCTGGA 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 TATATAACGTCATGGGGGAAGATGTGGCTTTCAGTACTTGGAGTGTATGAATGGTCTGGA 720

Qy       721 AATAATCCCCTGCCCCCTGAGATATGGCTCCTTCCATACATGCTTCCATTTCATCCAGGA 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 AATAATCCCCTGCCCCCTGAGATATGGCTCCTTCCATACATGCTTCCATTTCATCCAGGA 780

Qy       781 AGGATGTGGTGTCACTGCCGGATGGTCTATTTGCCGATGTCCTACTTATATGGCAAGAGG 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 AGGATGTGGTGTCACTGCCGGATGGTCTATTTGCCGATGTCCTACTTATATGGCAAGAGG 840

Qy       841 TTTGTTGGTCCAATCTCACCAACAGTATTATCTTTGAGAAAAGAGCTTTATACAGTACCA 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 TTTGTTGGTCCAATCTCACCAACAGTATTATCTTTGAGAAAAGAGCTTTATACAGTACCA 900

Qy       901 TACCATGATATAGATTGGGATCAGGCTCGCAATTTGTGTGCAAAGGAAGATTTGTACTAT 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 TACCATGATATAGATTGGGATCAGGCTCGCAATTTGTGTGCAAAGGAAGATTTGTACTAT 960

Qy       961 CCTCACCCACTTGTACAGGATATTCTTTGGGCATCTCTACACAAGTTCCTTGAGCCTATT 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       961 CCTCACCCACTTGTACAGGATATTCTTTGGGCATCTCTACACAAGTTCCTTGAGCCTATT 1020

Qy      1021 CTGATGCATTGGCCTGGAAAAAGATTGAGGGAAAAGGCTATTATTTCTGCATTGGAGCAT 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1021 CTGATGCATTGGCCTGGAAAAAGATTGAGGGAAAAGGCTATTATTTCTGCATTGGAGCAT 1080

Qy      1081 ATACATTACGAAGATGAGAATACTCGATATATTTGCATAGGTCCTGTAAATAAGGTGTTA 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1081 ATACATTACGAAGATGAGAATACTCGATATATTTGCATAGGTCCTGTAAATAAGGTGTTA 1140

Qy      1141 AATATGCTTTGCTGTTGGGTGGAAGATCCAAATTCTGAGGCCTTCAAGTTGCATCTTCCC 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1141 AATATGCTTTGCTGTTGGGTGGAAGATCCAAATTCTGAGGCCTTCAAGTTGCATCTTCCC 1200

Qy      1201 AGGATTTATGATTATCTATGGATTGCAGAAGATGGCATGAAAATGCAGGGCTACAATGGA 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1201 AGGATTTATGATTATCTATGGATTGCAGAAGATGGCATGAAAATGCAGGGCTACAATGGA 1260

Qy      1261 AGTCAACTATGGGACACTGCTTTTGCTGTCCAAGCAATTATTGCATCTAACCTCATTGAA 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1261 AGTCAACTATGGGACACTGCTTTTGCTGTCCAAGCAATTATTGCATCTAACCTCATTGAA 1320

Qy      1321 GAATTTGGTCCAACTATAAGAAAAGCTCATACCTATATTAAGAATTCACAGGTTTTAGAA 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1321 GAATTTGGTCCAACTATAAGAAAAGCTCATACCTATATTAAGAATTCACAGGTTTTAGAA 1380

Qy      1381 GATTGTCCAGGTGATCTTAATAAATGGTACCGTCACATTTCAAAAGGTGCTTGGCCTTTT 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1381 GATTGTCCAGGTGATCTTAATAAATGGTACCGTCACATTTCAAAAGGTGCTTGGCCTTTT 1440

Qy      1441 TCAACTGGAGATCATGGATGGCCAATTTCTGACTGCACAGCTGAAGGACTGAAAGCTGTT 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1441 TCAACTGGAGATCATGGATGGCCAATTTCTGACTGCACAGCTGAAGGACTGAAAGCTGTT 1500

Qy      1501 CTATTACTATCCAAAATTGCACCAGAAATAGTTGGTGAGCCAATAGACGTGAAGCGATTA 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1501 CTATTACTATCCAAAATTGCACCAGAAATAGTTGGTGAGCCAATAGACGTGAAGCGATTA 1560

Qy      1561 TATGATTCTGTAAATGTCATTCTCTCACTACAGAATGAAGATGGTGGTTTTGCAACATAT 1620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1561 TATGATTCTGTAAATGTCATTCTCTCACTACAGAATGAAGATGGTGGTTTTGCAACATAT 1620

Qy      1621 GAGCTTAAACGATCTTATAATTGGTTGGAGATAATCAATCCTGCTGAAACTTTTGGTGAC 1680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1621 GAGCTTAAACGATCTTATAATTGGTTGGAGATAATCAATCCTGCTGAAACTTTTGGTGAC 1680

Qy      1681 ATCGTTATTGATTATCCTTATGTGGAATGTACATCAGCAGCGATTCAAGCTTTGGCATCA 1740
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1681 ATCGTTATTGATTATCCTTATGTGGAATGTACATCAGCAGCGATTCAAGCTTTGGCATCA 1740

Qy      1741 TTTAGGAAATTATATCCTGGGCATCGCCGAGAAGAAATACAACATTGTATCGATAAAGCC 1800
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1741 TTTAGGAAATTATATCCTGGGCATCGCCGAGAAGAAATACAACATTGTATCGATAAAGCC 1800

Qy      1801 ACTACCTTCATTGAAAAAATACAAGCTTCAGATGGATCATGGTATGGTTCTTGGGGAGTT 1860
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1801 ACTACCTTCATTGAAAAAATACAAGCTTCAGATGGATCATGGTATGGTTCTTGGGGAGTT 1860

Qy      1861 TGCTTCACTTACGGTGCTTGGTTTGGGGTAAAAGGTCTGATTGCTGCTGGAAGGAGTTTC 1920
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1861 TGCTTCACTTACGGTGCTTGGTTTGGGGTAAAAGGTCTGATTGCTGCTGGAAGGAGTTTC 1920

Qy      1921 AGTAATTGCTCAAGCATCCGTAAAGCTTGTGAATTTCTGCTGTCCAAGCAGCTTCCTTCT 1980
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1921 AGTAATTGCTCAAGCATCCGTAAAGCTTGTGAATTTCTGCTGTCCAAGCAGCTTCCTTCT 1980

Qy      1981 GGTGGCTGGGGAGAGAGTTATCTGTCCTGTCAAAACAAGGTGTATTCAAATCTGGAAGGC 2040
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1981 GGTGGCTGGGGAGAGAGTTATCTGTCCTGTCAAAACAAGGTGTATTCAAATCTGGAAGGC 2040

Qy      2041 AACAGGTCTCATGTGGTCAACACTGGGTGGGCTATGTTGGCTCTCATTGATGCTGGACAG 2100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2041 AACAGGTCTCATGTGGTCAACACTGGGTGGGCTATGTTGGCTCTCATTGATGCTGGACAG 2100

Qy      2101 GCTAAGAGAGATTCGCAACCACTGCACCGGGCAGCTGCATACTTGATAAATTCCCAATTG 2160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2101 GCTAAGAGAGATTCGCAACCACTGCACCGGGCAGCTGCATACTTGATAAATTCCCAATTG 2160

Qy      2161 GAGGACGGTGACTTTCCGCAGCAGGAAATAATGGGAGTCTTCAACAAGAATTGCATGATC 2220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2161 GAGGACGGTGACTTTCCGCAGCAGGAAATAATGGGAGTCTTCAACAAGAATTGCATGATC 2220

Qy      2221 ACATACGCCGCATACAGAAACATATTCCCCATTTGGGCGTTGGGAGAATACCAATCCCAA 2280
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2221 ACATACGCCGCATACAGAAACATATTCCCCATTTGGGCGTTGGGAGAATACCAATCCCAA 2280

Qy      2281 GTATTGCAATCTCGTTAATCGAGCCTTAGTTAGGGTGTCATCCTAATTATTTCGACCTGT 2340
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2281 GTATTGCAATCTCGTTAATCGAGCCTTAGTTAGGGTGTCATCCTAATTATTTCGACCTGT 2340

Qy      2341 GGACAGTAAAAGAAATAATAATAATAACCTATTTTTCTTTTATTTCCATGGCTCTCTTAA 2400
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2341 GGACAGTAAAAGAAATAATAATAATAACCTATTTTTCTTTTATTTCCATGGCTCTCTTAA 2400

Qy      2401 AATGTTTGTGACTAATGAGTTTAGTAGTCAGCTAAAAAAAAAAAGCAAACACGTGGAGAAT 2460
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2401 AATGTTTGTGACTAATGAGTTTAGTAGTCAGCTAAAAAAAAAAAGCAAACACGTGGAGAAT 2460

Qy      2461 GCCTGTAAGTTTTTTCTATTACTCATAGACGCCTCTCTTGCTTTCCCTGCAGCAGAGGAA 2520
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2005, 06:30:52 ; Search time 419 Seconds
(without alignments)
9997.299 Million cell updates/sec

Title: US-09-419-679-5
Perfect score: 2560
Sequence: 1 ttggcctcttgccagcaaaa.....agatataaaaaaaaaaaaaa 2560

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

other seqs from Parent Appl's

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2560	100.0	2560	4	US-09-419-679-5	Sequence 5, Appli
2	1183.6	46.2	2558	4	US-09-419-679-1	Sequence 1, Appli
3	944	36.9	1882	4	US-09-419-679-3	Sequence 3, Appli
4	563.8	22.0	1300	4	US-09-419-679-7	Sequence 7, Appli
5	304.8	11.9	4346	4	US-09-919-039-11[illegible]	Sequence 112, App
6	280.2	10.9	4239	4	US-09-815-048-1	Sequence 1, Appli
7	211	8.2	2196	4	US-09-614-221A-308	Sequence 308, App
8	74.4	2.9	747	4	US-09-248-796A-3811	Sequence 3811, Ap
c 9	70.4	2.8	539	3	US-08-998-416-229	Sequence 229, App
10	58.2	2.3	173	4	US-09-313-294A-804	Sequence 804, App
c 11	46	1.8	1141	4	US-09-806-708B-22	Sequence 22, Appl
c 12	44	1.7	1664976	4	US-08-916-421B-1	Sequence 1, Appli
c 13	44	1.7	1664976	4	US-09-692-570-1	Sequence 1, Appli
14	43.6	1.7	403	4	US-08-956-171B-767	Sequence 767, App
15	43.6	1.7	403	4	US-08-781-986A-767	Sequence 767, App
16	43	1.7	1141	4	US-09-806-708B-22	Sequence 22, Appl
17	42.8	1.7	63183	4	US-09-949-016-13047	Sequence 13047, A
18	42.8	1.7	63183	4	US-09-949-016-13048	Sequence 13048, A
19	42.6	1.7	3091	4	US-09-949-016-3440	Sequence 3440, Ap
c 20	42.6	1.7	3217	4	US-09-949-016-2400	Sequence 2400, Ap
c 21	42.6	1.7	3236	4	US-09-949-016-2399	Sequence 2399, Ap
c 22	42.6	1.7	3463	4	US-09-949-016-1417	Sequence 1417, Ap
c 23	42.6	1.7	4287	1	US-08-244-189-1	Sequence 1, Appli
c 24	42.6	1.7	4287	1	US-08-306-691B-53	Sequence 53, Appl
25	42.6	1.7	62474	4	US-09-949-016-15182	Sequence 15182, A
c 26	42.6	1.7	108060	4	US-09-949-016-13159	Sequence 13159, A
c 27	42.6	1.7	121234	4	US-09-949-016-14142	Sequence 14142, A
c 28	42.6	1.7	140315	4	US-09-949-016-14141	Sequence 14141, A
c 29	41.2	1.6	7218	1	US-08-232-463-14	Sequence 14, Appl
30	40.4	1.6	255	4	US-09-248-796A-3810	Sequence 3810, Ap
31	40.4	1.6	1599	3	US-09-277-565-16	Sequence 16, Appl
32	40.4	1.6	1830121	4	US-09-557-884-1	Sequence 1, Appli
33	40.4	1.6	1830121	4	US-09-643-990A-1	Sequence 1, Appli
34	40	1.6	5319	2	US-08-861-464-7	Sequence 7, Appli
35	40	1.6	5319	2	US-08-396-001-7	Sequence 7, Appli
36	40	1.6	5319	3	US-09-323-433A-7	Sequence 7, Appli
37	40	1.6	5319	4	US-09-826-752-7	Sequence 7, Appli
c 38	40	1.6	640681	4	US-09-790-988-1	Sequence 1, Appli
c 39	39.8	1.6	3081	3	US-09-319-989-9	Sequence 9, Appli
c 40	39.8	1.6	3360	3	US-09-319-989-8	Sequence 8, Appli
41	39.8	1.6	42450	4	US-09-815-048-3	Sequence 3, Appli
c 42	38.6	1.5	601	4	US-09-949-016-70991	Sequence 70991, A
c 43	38.6	1.5	276687	4	US-09-949-016-13840	Sequence 13840, A
c 44	38.4	1.5	8534	4	US-09-949-016-14723	Sequence 14723, A
c 45	38.4	1.5	148156	4	US-09-949-016-11776	Sequence 11776, A

ALIGNMENTS

RESULT 1
US-09-419-679-5
; Sequence 5, Application US/09419679
; Patent No. 6630617
; GENERAL INFORMATION:
; APPLICANT: Falco, S. Carl
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Hitz, William D.
; APPLICANT: Kinney, Anthony J.
; APPLICANT: Rafalski, Antoni
; APPLICANT: McGonigle, Brian
; APPLICANT: Lohman, Karin
; TITLE OF INVENTION: Enzymes Involved In Squalene Metabolizm
; FILE REFERENCE: BB1113 US NA
; CURRENT APPLICATION NUMBER: US/09/419,679
; CURRENT FILING DATE: 1999-10-15
; EARLIER APPLICATION NUMBER: 60/105,405
; EARLIER FILING DATE: 1998-10-23
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 5
; LENGTH: 2560
; TYPE: DNA
; ORGANISM: Glycine max
US-09-419-679-5

Query Match 100.0%; Score 2560; DB 4; Length 2560;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2560; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 TTGGCCTCTTGCCAGCAAAACAGAATGTGGAAGCTCAAGTTCGCCGAAGGAGGGAATCCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TTGGCCTCTTGCCAGCAAAACAGAATGTGGAAGCTCAAGTTCGCCGAAGGAGGGAATCCA 60

Qy         61 TGGCTTCGGACATTGAACAATCACGTTGGAAGACAGGTGTGGGAGTTCGATCCTAAGCTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGGCTTCGGACATTGAACAATCACGTTGGAAGACAGGTGTGGGAGTTCGATCCTAAGCTT 120

Qy        121 GGATCGCCGCAAGATCTCCTCGAGATTGAGAAAGCTCGCCAGAATTTTCACGATAACCGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATCGCCGCAAGATCTCCTCGAGATTGAGAAAGCTCGCCAGAATTTTCACGATAACCGC 180

Qy        181 TTTACCCACAAACACAGCGCTGATCTACTTATGCGGATGCAGTTCGCAAGAGAGAACCCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTTACCCACAAACACAGCGCTGATCTACTTATGCGGATGCAGTTCGCAAGAGAGAACCCA 240

Qy        241 ACACGTGAAGTCTTGCCCAAAGTCGGAGTTAAGGATATTGAGGATGTGACCCAAGAGATT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACACGTGAAGTCTTGCCCAAAGTCGGAGTTAAGGATATTGAGGATGTGACCCAAGAGATT 300
```